Stochastic State Space Modelling of Nonlinear systems - With application to Marine Ecosystems

Møller, Jan Kloppenborg

Publication date:
2011

Document Version
Publisher's PDF, also known as Version of record

Stochastic State Space Modelling of Nonlinear systems

- With application to Marine Ecosystems

Jan Kloppenborg Møller

Kongens Lyngby 2010
IMM-PHD-2010-246

Technical University of Denmark
Informatics and Mathematical Modelling
Building 321, DK-2800 Kongens Lyngby, Denmark
Phone +45 45253351, Fax +45 45882673
reception@imm.dtu.dk
www.imm.dtu.dk

National Environmental Research Institute
Aarshus University
Fredriksborgvej 399, DK-4000 Roskilde, Denmark
Phone +45 46301200, Fax: +45 46301114
dmu@dmu.dk
www.dmu.dk

IMM-PHD: ISSN 0909-3192

Preface

This thesis was prepared at Informatics and Mathematical Modelling, the Technical University of Denmark in partial fulfilment of the requirements for acquiring the Ph.D. degree in engineering.

The main topic of the thesis is stochastic dynamical models for marine ecosystems. Both discrete and continuous time models are considered, and also more general aspects of the methods are considered. The thesis consists of a summary report and a collection of five research papers written during the period 2006–2010, and submitted to international journals.

I would like to thank my supervisors Prof. Jacob Carstensen (NERI) and Prof. Henrik Madsen (DTU-Informatics) for support and valuable discussions throughout the project period. During the spring and summer of 2007 I visited "Biologisk Institut" at the University of Oslo and I would like to thank the staff and in particular Prof. Tom Andersen for their hospitality during my stay there. The project was funded by the EU-project "Thresholds" (though NERI), the Danish Graduate School in Biostatistic and DTU, and I would like to acknowledge all institutions for their financial support.

Lyngby, December 2010

Jan Kloppenborg Møller

Abstract

This thesis deals with stochastic dynamical systems in discrete and continuous time. Traditionally dynamical systems in continuous time are modelled using Ordinary Differential Equations (ODEs). Even the most complex system of ODEs will not be able to capture every detail of a complex system like a natural ecosystem, and hence residual variation between the model and observations will always remain. In stochastic state-space models the residual variation is separated into observation and system noise and a main theme of the thesis is a proper description of the system noise. Additive Gaussian noise is the standard approach to introduce system noise, but this may lead to undesirable consequences for the state variables. In biological models, where the state-space generally contains positive real numbers only, modelling in the log-domain ensures positive state variables, however, this transformation is likely to conflict with the concept of mass balances. One of the central conclusions of the thesis is that the stochastic formulations should be an integral part of the model formulation.

As discrete-time stochastic processes are simpler to handle numerically than continuous-time stochastic processes, I start by considering discrete-time processes. A novel approach combining multiplicative and additive log-normal noise has been developed in discrete time, and used to demonstrate the effect of stochastic forcing in simple discrete-time regime shift models. An approximate maximum likelihood estimation procedure based on the second order moment representation of the multiplicative and additive log-normal noise model was developed and tested in simulation studies.

The transition to continuous-time stochastic models (here Stochastic Differential Equations (SDEs)) offers the opportunity of embedding parts of the ODE

processes into the stochastic part of the model (the diffusion term). The estimation method we use here (maximum likelihood and the Extended Kalman Filter (EKF)) rely on state-independent diffusion, but for a wide class of SDEs there exist an alternative description (given by the Lamperti transform) of the input-output relation, where the diffusion term is independent of the state. This alternative description is used to develop better parametric descriptions of the diffusion term, while maintaining the opportunity of estimation by standard software.

Additionally, the state-space formulation facilitates estimation of unobserved states. Based on estimation of random walk hidden states and examination of simulated distributions and stationarity characteristics, a methodological framework for structural identification based on information embedded in the observations of the system has been developed. The applicability of the methodology is demonstrated using phytoplankton and nitrogen data from a Danish estuary as well as bacterial growth data from a controlled experiment.

In summary, the novelty of the work presented here is the introduction of more appropriate stochastic descriptions in non-linear state-space models, which can include combinations of additive and multiplicative noise components under various distributional assumptions. A model identification and estimation framework for working with such models has been developed and tested using data from biological and ecological systems typically characterised by non-linear and non-Gaussian responses.

Resumé

(Abstract in Danish)

Denne afhandling beskæftiger sig med stokastiske dynamiske systemer i såvel diskret som kontinuert tid. Traditionelt modelleres kontinuert tids dynamiske systemer med ordinære differentialligninger (ODEer). Selv det mest komplekse system af ODEer vil ikke være i stand til at beskrive alle detaljer af et komplekst system som eksempelvis et naturligt økosystem, og der vil derfor altid være en residual variation mellem modellen og observationer af systemet. I stokastiske tilstands modeller separeres residual variationen i observations- og systemstøj og et hovedtema i afhandlingen er en korrekt beskrivelse af system støjen. Standard tilgangen til at introducere system støj er additiv Gaussisk støj, denne tilgang kan dog føre til uønskede konsekvenser for tilstandsvariablene. I biologiske modeller, hvor tilstandsrummet generelt kun indeholder positive reelle tal, vil modellering i log-domænet sikre positive tilstandsvariable, log-transformationen vil dog i mange tilfælde være i konflikt med massebalance konceptet. En af de centrale konklusioner i afhandlinger er, at den stokastiske formulering bør være en integreret del af modelformuleringen.

Diskret tids modeller er simplere at håndtere numerisk end kontinuert tids modeller og jeg starter derfor med at se på diskret tids modeller. En ny tilgang der kombinerer multiplikativ og additiv log-normal støj er blevet udviklet og benyttet til at demonstrerer stokastisk belastning i simple regime skift modeller. Baseret på anden orden moments repræsentationen af den multiplikative og additive log-normal støjmodel er en approksimativ maksimum likelihood metode udviklet og testet i simulerings studier.

Overgangen til kontinuert tids stokastiske modeller (her stokastiske differential ligninger (SDEer)) giver mulighed for at indlejre dele of ODE-processen i den stokastiske del af modellen (diffusions leddet). Den estimations metode vi bruger her (maksimum likelihood og det Udvidede Kalman Filter (EKF)) er baseret på tilstandsuafhængig diffusion, men for en stor klasse af SDEer eksisterer der en alternativ beskrivelse (givet ved Lamperti transformationen) af relationen mellem input og output, hvor diffusionsleddet er tilstandsuafhængigt. Denne alternative beskrivelse bruges til at udvikle bedre parametriske beskrivelser af diffusionsleddet, mens muligheden for estimering i standard programmer bibeholdes.

Tilstandsrum repræsentationen giver mulighed for estimering af ikke observerede tilstande. Baseret på estimering af ikke observerede "random walk" (tilfældig gang) tilstande og analyse af simulationsfordelinger og stationære karakteristika er en metodisk ramme for strukturel identifikation baseret på information indlejret i observationerne blevet udviklet. Anvendelsen af metoden er demonstreret med fytoplankton og nitrogen data fra en dansk fjord og bakterievækst data fra et kontrolleret eksperiment.

Sammenfattende er det nyskabende i det præsenterede arbejde introduktionen af en bedre stokastisk beskrivelse i ikke lineære tilstandsrums modeller, der kan inkludere kombinationer af additive of multiplikative støj komponenter under forskellige fordelingsantagelser. En model identifikation og estimerings ramme for udvikling af sådanne modeller er blevet udviklet og testet med data fra biologiske og økologiske system der typisk karakteriseres af ikke lineær og ikke Gaussisk dynamik.

List of publications

Papers included in the thesis

[A] Jan Kloppenborg Møller, Jacob Carstensen, Henrik Madsen, and Tom Andersen. Dynamic two state stochastic models for ecological regime shifts. *Environmetrics*, **20**, 912-927, (2009).

[B] Jan Kloppenborg Møller, Henrik Madsen, and Jacob Carstensen. Kalman Filtering and Parameter Estimation in State Space Models with Multiplicative Noise. *Submitted to IEEE Transactions on Signal Processing*, (2010).

[C] Jan Kloppenborg Møller, Henrik Madsen, and Jacob Carstensen. Parameter Estimation in a Simple Stochastic Differential Equation for Phytoplankton Modelling. *Ecological Modelling*, **222**, 1793-1799, (2011).

[D] Jan Kloppenborg Møller, and Henrik Madsen. From Level Dependent Diffusion to Constant Diffusion in Stochastic Differential Equations by the Lamperti Transform. *IMM-technical report 2010-16*, (2010).

[E] Jan Kloppenborg Møller, and Henrik Madsen. Applying EKF for Maximum Likelihood Estimation in Stochastic Differential Equations with State Dependent Diffusion. *To be submitted for Automatica*, (2010).

[F] Jan Kloppenborg Møller, Henrik Madsen, and Jacob Carstensen. Structural identification and validation in stochastic differential equation based models - With application to a marine ecosystem NP-model. *Submitted to Journal of the Royal Statistical Society, Series C*, (2010).

[G] Jan Kloppenborg Møller, Kirsten Riber Phillipsen, Lasse Engbo Christiansen, and Henrik Madsen. Development of a Restricted State Space Stochastic Differential Equation Model for Bacterial Growth in Rich Media *Submitted to Journal of Theoritical Biology*, (2010).

Other publications

The following papers were also prepared during the project period. The scientific content is covered or partly covered by the included papers, thus these papers are not included.

- Møller J.K., Carstensen J., and Madsen H. "Regime Shift Models for Simulation of the Interaction between Benthic and Pelagic Production ". *IMM-technical report 2006-23*, (2006).
- Møller J.K., Carstensen J., and Madsen H. "Parameter Estimation in State Space Models with Multiplicative Noise - Examples". *IMM-technical report 2009-06*, (2009).
- Møller J.K., Carstensen J., Madsen H., and Andersen T. "Stochastic dynamical models for ecological regime shifts". Poster for *ASLO Aquatic Sciences meeting*, Nice, France (2009).
- Møller J.K., Carstensen J., and Madsen H. "Identification of ecosystem parameters by SDE-modelling". Contributed session at *Annual conference of the International Environmetrics Society.* Margarita Island, Venezuela, (2010).
- Møller J.K., Christiansen L.E., and Madsen H. "Stochastic Dynamical Systems". Part II of "An introduction to Mathematical Modelling, Nonlinear, Stochastic and Complex Systems", MMC summer school (2010).

The following papers were partly prepared during the project period. The scientific content is however out of the scope of this thesis and these are therefore not included in the thesis.

- Møller J.K., Madsen H., and Nielsen H.Aa. "Algorithms for Adaptive Quantile Regression - and a Matlab Implementation". *IMM-technical report 2006-08*, (2006)

- Pinson P., Nielsen H.Aa., Møller J.K., Madsen H., and Kariniotakis G.N. "Non-parametric probabilistic forecasts of wind power: required properties and evaluation". *Wind Energy*, **10**(6), 497-587, (2007).
- Pinson P., Møller J.K., Nielsen H.Aa., Madsen H., and Kariniotakis G.N. "Evaluation of Nonparametric Probabilistic Forecasts of Wind Power". *MM-Technical Report-2007-02*, (2007).
- Møller J.K., Nielsen H.Aa., and Madsen H. "Time-adaptive quantile regression". *Computational Statistics & Data Analysis*, **52**(3), 1292-1303 (2008).
- Møller J.K., Madsen H., and Nielsen H.Aa. "Combined Forecast and Quantile Regression for Wind Power Prediction.". *IMM-technical report 2007-19*, (2008).

Contents

Part I

Summary report

Chapter 1

Introduction

Ecosystems and marine ecosystems in particular are complex systems of interconnected processes evolving in time and space. A detailed understanding of the individual subprocesses paves the way for complex deterministic modelling by ordinary differential equations (ODEs) and even partial differential equations (PDEs). This strategy implies increasing complexity of ecosystem models, and inclusion of more states as process understanding increases. As a consequence, an increasing number of parameters needs to be included and estimated in ecosystem models.

In addition to the evolving process understanding, the concept of alternative stable states in marine ecosystems has been recognised since the 1960's (Schindler, 2006). Even though recognised for a long time, the mechanistic understanding of the underlying mechanisms are still at an early stage and described using simple ecosystem models. Simple ecosystem models do not attempt to include detailed interactions into the models and unknown and unobserved (or known but not included) effects needs to be accounted for. This can be done by including stochastic noise working within the system. The inclusion of stochastic noise is natural in the sense that a complete reductionistic description of an open and complex system like natural ecosystems is not realistic. Stochastic perturbations introduce transitions probabilities on the system, which in turn imply that we can assign a likelihood to a given model. This enables statistical testing and quantification of modelling results.

The analysis of non-linear phenomena, like e.g. regime shifts, can follow different strategies. One is a strictly statistical approach, where changes in the linear dynamics of the system are analysed to identify changes in the linear dynamics of the system (e.g. change point detection). Another strategy is to include non-linear mechanisms in a process-based description of the system and analyse how this affects the dynamics of the system. The latter is the strategy followed in the present work.

In deterministic systems, the solution of a dynamic system is known (for complex systems by numerical solutions), when the initial conditions are known. The transition to alternative stable states will be determined solely by the forcing imposed on the system. In stochastic systems this is replaced by a probability of switching to an alternative state given that such a state exist. This also implies that we will only be able to assign a transition probability, not a certain threshold.

A key issue in stochastic modelling is parameter and state estimation. Parameter estimation deals with inference of parameters such that, given a model structure, the best description of data is obtained in some sense. In this thesis we use the maximum likelihood sense. State estimation deals with finding the most probable (or expected) path of the state given the information provided by the observations of the system. In the context of marine ecosystem modelling both issues are relevant. While parameter estimation answers questions on ecosystem functioning, state estimates can answer questions about the past states of the system.

The reasoning of mass flows in ecosystems is most convenient in a differential setting, because statements like "*species A eats B but not C and B eats C*", translate directly into mass flows in the corresponding differential equations. In contrast a discrete time formulation might need to include mass flow from C to A. The mathematical setup is however simpler in a discrete time setting, because e.g. transition probabilities are easier to calculate. Chapter 2 deals with the conceptual idea of regime shifts in a discrete stochastic time setting, and a multiplicative-additive log-normal noise model is introduced. Chapter 3 deals with state and parameter estimations in discrete time, and in particular a likelihood based estimation procedure of the additive-multiplicative noise model is presented.

In the same way as mass flows evolve in continuous time, it is reasonable to assume that random perturbations of an ecosystem also propagates through the system in continuous time. This leads to stochastic differential equations (SDEs). One of the key advantages of SDE formulations is that natural constraints are (in principle) easily built into the differential noise term (referred to as the diffusion term). The conceptual understanding of ecosystems is better

described in continuous time. Therefore model development based on a combination of inspection of state estimates and hypothesis based on literature is more straightforward than in discrete time. Chapter 4 deals with state and parameter estimation in SDEs, and in particular a method for combining the EKF and state dependt diffusion is discussed, further a methodological framework for pinpointing model deficiencies and inferring structural improvements based on embedded information in the observations is presented.

The objective of the summary report given in Chapter 2-5 is to present the main results of the studies presented in papers A-G. Furthermore the summary report presents some of the underlying theory and provides some considerations that are not covered in the papers, and Chapter 5 provides an overall conclusion and discussion of the presented results.

Chapter 2

Discrete time modelling

The aim of a dynamical model is often to predict the future or to describe events of the past. In ecosystem modelling these aims are all highly relevant. While good predictions are important for managements strategies, good descriptions of the past based on the available information are important for understanding ecosystem functioning.

An often hypothesised threshold lies in the benthic-pelagic interaction. The pelagic plants (primarily phytoplankton) have a short lifespan and their growth mostly relies on nutrient availability in the water-column. Due to the mixing effects pelagic plants are more efficient in utilising solar radiation for photosynthesis. On the other hand, the benthic vegetation (in Danish waters primarily eelgrass) has a longer lifespan and uses nutrients in the sediment for growth. During the past century the depth limit for eelgrass has significantly decreased. While there can be many reasons for this decrease in depth limit, one of the widely recognised effects is the outshading from phytoplankton. The hypothesis is that increasing phytoplankton biomass (as a result of increasing nutrient loading) reduced the amount of sunlight reaching the benthic region and therefore photosynthetic growth of the benthic vegetation. This may lead to enhanced nutrient releases from the sediments sustaining increased phytoplankton growth.

This situation can potentially lead to hysteresis effects and substantial reductions in nutrient input will be needed to restore a past and desired state. In the

worst case scenario recovery may not be possible in practice, since reduction of pelagic biomass production alone cannot restore the ecosystem. Bistability of ecosystems have been analysed on the conceptual level for many years (Schindler, 2006) with Holling (1973) and May (1977) as early examples, while Carpenter et al. (1999) and Carpenter (2005) are more recent examples of conceptual models for regime-shift in aquatic ecosystems. While these studies are theoretical studies of conceptual regime-shift models, empirical evidence is rather sparse, one example is Scheffer et al. (2003), where the existence of bistability is shown in a controlled experiment.

This chapter deals with regime shift models in discrete time. Selected results from Paper A are presented in Section 2.4, while Sections 2.1-2.3 present the development which leads to the formulation of the models presented in Paper A.

2.1 Simple regime shift models

One of the most simple regime shift models one can think of is the Self-Exciting Threshold Moving Average (SETAR) model (Tong, 1990) of order one

$$X_t = a_0^{(j)} + a_1^{(j)} X_{t-1} + h^{(j)} \epsilon_t, \tag{2.1}$$

where $a_0^{(j)}$, $a_1^{(j)}$, and $h^{(j)}$ are constants depending on which regime (R_j, $j \in \{1, ..., K\}$) X_{t-1} belongs to, and ϵ_t is a white noise process. The dynamics of the system (2.1) is driven by the state itself, and fixing $h^{(j)} = 0$ gives the deterministic skeleton of the model. Analysis of the deterministic part of the model gives some hints about the expected dynamics. The stochastic part does, however, also contribute to the dynamics of the system. Random behaviour is expected in marine ecosystem, where the dynamics of the system are governed by very complex interactions of interconnected processes, which are often only partly observed. Further, weather conditions like wind and rainfall influence the system strongly.

The SETAR model is illustrative for regime-shift models, because the threshold is explicitly formulated and the dynamics of the linear model change as the state variable (X_t) exceeds a certain threshold. In the most simple setup with only two regimes, the SETAR model (2.1) can be written as

$$X_t = \begin{cases} a_0^{(1)} + a_1^{(1)} X_{t-1} + h^{(1)} \epsilon_t, & X_{t-1} \leq r \\ a_0^{(2)} + a_1^{(2)} X_{t-1} + h^{(2)} \epsilon_t, & X_{t-1} > r \end{cases}. \tag{2.2}$$

The dynamics changes when (X_t) exceeds the threshold (r). Even though the simplicity of the SETAR model is appealing the function is not differentiable at the threshold point and it can therefore not represent a discrete-time equivalent of a differential equation. A more flexible class of models is the first order Smooth Threshold Autoregressive (STAR) model (Tong, 1990), defined as

$$X_t = a_0 + a_1 X_{t-1} + (b_0 + b_1 X_{t-1})\, f(X_{t-1}), \tag{2.3}$$

where $f(x)$ is a continuously increasing function of x. The dynamics of (2.3) fits better with the conceptual understanding of regime shift in ecology, because f can be given a biological or physical interpretation, like Michaelis-Menten (or Monod) kinetics for process rates

$$f(x) = \frac{x}{c + x}, \tag{2.4}$$

where c is called the half-saturation constant. Carpenter (2005) suggests a generalisation of (2.4)

$$f(x) = \frac{x^q}{c + x^q}, \tag{2.5}$$

to describe the smoothness (in an ODE setting) of the threshold in a simple regime-shift model based on eutrophication in an aquatic ecosystem (the smoothness is described by q).

2.2 A benthic-pelagic interaction model

Even though the one-dimensional models presented above help us to understand the nature of threshold models, the interesting conclusions form when we are able to infer consequences about the interaction between ecosystem components. For this purpose the one-dimensional models will not suffice and a_i, b_i, and f needs to be replaced by matrix/vector functions.

The conceptual setting for a regime-shift model of benthic-pelagic interaction is presented in Figure 2.1. When nutrient loading is low phytoplankton production is also low and sunlight is able to reach the benthic region and sustain photo-synthetic growth of the rooted vegetation. As nutrient loading to marine ecosystems (primarily nitrogen as the most limiting factor) increases, phytoplankton production also increases preventing sunlight from reaching the benthic region. Particulate organic material resulting from phytoplankton growth reaches the sediments, where it is permanently buried, used for benthic fauna growth or recycled back into the water column. The benthic-pelagic interaction

and the fate of the sedimenting organic material depend, among other things, on the stability of the sediments. High coverage of benthic vegetation helps to stabilise the sediments, while low coverage of benthic vegetation will not be able to stabilise the sediments and consequently more nutrients will be recycled from the sediments to the water-column. A mathematical formulation of this feed-back may potentially lead to a regime-shift model.

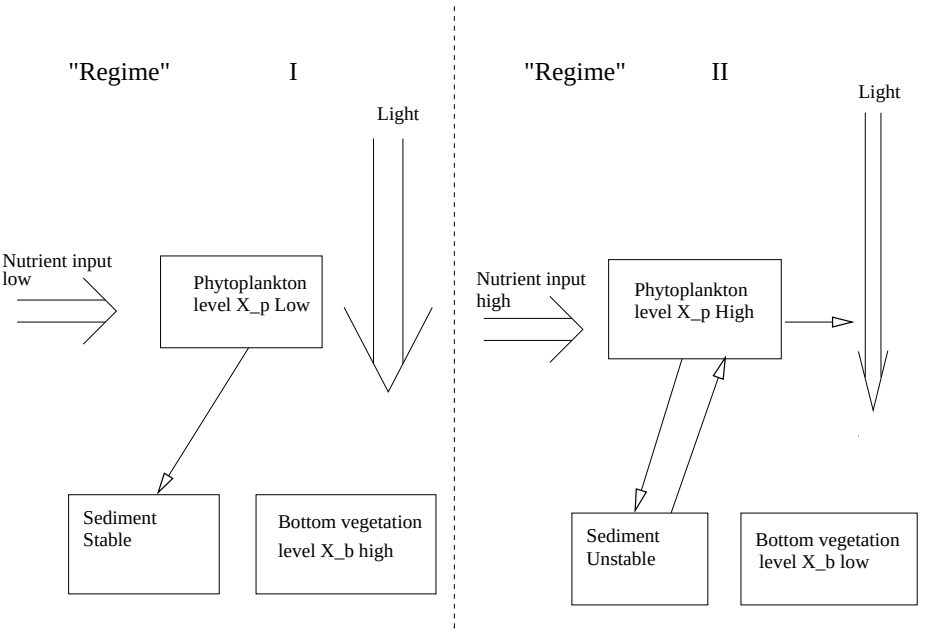

Figure 2.1: Conceptual diagram of a simple regime shift model for benthic pelagic interaction. At low nutrient inputs the pelagic production is low and the water clear, implying that sunlight can reach the benthic region and sustain photo synthetic growth of benthic vegetation, stabilising the sediments, and enhancing permanent burial of nutrients. The pelagic production increases with nutrient inputs, which outshades benthic vegetation and enhances nutrient releases from the sediment, causing a negative feedback.

A simple mathematical formulation of the setup in Figure 2.1 can be given like

$$\begin{bmatrix} X_{p,t} \\ X_{b,t} \end{bmatrix} = \begin{bmatrix} a_p N_{ex,t} \\ kf(X_{p,t}) \end{bmatrix} + \begin{bmatrix} b_p^{(j)} & 0 \\ b_{pb} f(X_{p,t}) & b_b \end{bmatrix} \begin{bmatrix} X_{p,t-1} \\ X_{b,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{p,t} \\ \epsilon_{b,t} \end{bmatrix}, \quad (2.6)$$

with

$$b_p^{(j)} = \begin{cases} t_p & \text{if } X_{b,t-1} > r_b \\ t_p + d & \text{if } X_{b,t-1} \leq r_b \end{cases}, \quad (2.7)$$

where $X_{p,t}$ and $X_{b,t}$ is the pelagic and benthic vegetation respectively (at time t), a_p is the part of nutrient loading which stays in the system, b_b is a survival factor for the benthic vegetation, $\boldsymbol{\epsilon}_t = [\epsilon_{p,t}, \epsilon_{b,t}]^T$ is a Gaussian white noise process, with covariance Σ, $b_p^{(j)}$ is a survival factor for phytoplankton including recycling and the value of $b_p^{(j)}$ depends on the coverage of the benthic vegetation, b_{pb} is the growth factor for benthic vegetation and is limited by $f(x)$ which describes the outshading stemming from phytoplankton. Finally $kf(x)$ describes the ability of seeds from outside the system to settle within the system. The function f should be a decreasing function of the phytoplankton biomass, and the one chosen here is

$$f(x; r_p, \tau_p) = 1 - \Phi(x; r_p, \tau_p^2), \tag{2.8}$$

where Φ is the Gaussian CDF with r_p controlling the location of the threshold and τ_p describe the smoothness of the threshold. The model (2.6) is a mixture between a hard threshold and a smooth threshold model, with $b_p^{(j)}$ describing a hard threshold and $f(\cdot)$ describing a smooth threshold. Clearly the hard threshold could be replaced by a smooth threshold. A simple formulation is to replace $b_p^{(j)}$ with

$$b_p(X_{b,t}) = (1 - \Phi(X_{b,1}; r_b, \tau_b^2))d + t_p, \tag{2.9}$$

where r_b and τ_b again describe the location and the smoothness of the threshold. The deterministic skeleton of the hard threshold (referred to as SETAR) and the smooth threshold (referred to as STAR) models are presented in Figure 2.2, where the deterministic models are very similar. For low loadings the SETAR model show a noisy nature, which are stable limit cycles, where the stable point for the SETAR model in each regime is a point in the other regime. This kind of behaviour is not present in the smooth threshold setup, and the span of loadings giving rise to stable limit cycles that can be derived analytically (Møller et al., 2006). However, these limit cycles are more a consequence of the mathematical formulation and are unlikeli to be observed in a real ecosystem. The main reason is that the hard threshold model cannot constitute a solution to an ODE.

The threshold model in Figure 2.2 has a large span of loading with alternative stable states (from about 1.5 to 5). If loading is increased above 5, the benthic vegetation becomes completely outshaded and loading must be reduced to less than 1.5 to restore a "healthy" system. We will refer to this span with two alternative stable states as the hysteresis span, and this varies with the chosen parameter values.

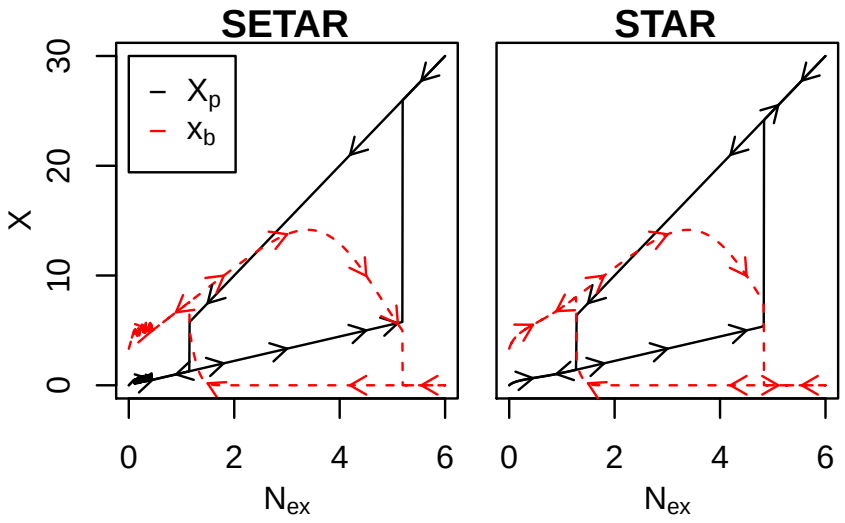

Figure 2.2: Hard (left column) and smooth (right column) threshold for the model (2.6) and (2.7) (hard threshold) and the model (2.6) and (2.9) (smooth threshold), with $a_p = 1$, $t_p = 0.1$, $d = 0.7$, $k = 1$, $b_{pb} = 1$, $\tau_p = \tau_b = 1$, $r_p = r_b = 5$. The two panels are very similar, besides the noisy appearance for low loadings in the SETAR model.

2.3 Adding noise

The analysis of the deterministic skeleton of a stochastic system gives information on the behaviour of the system under different loadings. To assess different management strategies it is necessary to include distributional properties of model. For example, if there is a positive economical gain from increased nitrogen loading and a negative economical gain from low benthic vegetation (or just a preference of the clear water situation), then the preferred strategy from the deterministic solution is to increase loading to just below the critical value. A real life system will be subjected to random perturbations and if the system is close to a threshold the risk of exceeding this threshold becomes large. The cost of restoring the system within a reasonable time horizon might be large compared to the cost of managing the system to a state with a small risk of ecosystem collapse.

The situation is exemplified in Figure 2.3, with the initial state of the system set as the stationary state of the healthy system with a external loading of 3. Even though there is a small drift towards the unhealthy situation most trajectories

stay in the healthy state until the mean loading is increased to 6 (at $t = 10$). With this loading there is a small chance only that the system remains in the healthy regime, while most of the simulated trajectories follow the path to the collapse of benthic vegetation. In an attempt to restore the previous healthy situation, the mean loading is substantially decreased to about 2. Even though, at this level, there is a chance that the ecosystem could recover, most trajectories still suggest a phytoplankton dominated system. A further reduction of the loading to 1 (outside the region of bistability), yields the desired management results and most trajectories end up in the healthy situation after 20 time steps of low loading. Having undergone a regime shift to the healthy state, increasing the loading to 3 (in mean), can be considered sustainable from a management point-of-view.

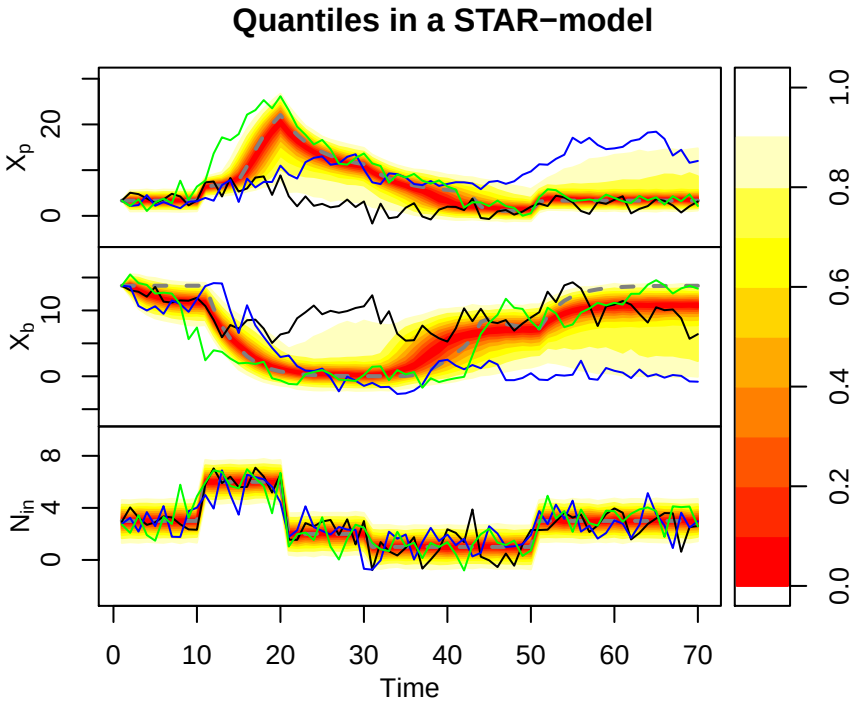

Figure 2.3: Simulation with the smooth threshold model presented in Figure 2.2 right column and parameter values for the deterministic skeleton as given in the caption of Figure 2.2. The stochastic noise is added as uncorrelated Gaussian noise in each state as well as in the input, all with variance equal 1. Colour code refer to quantiles in the simulated distribution, coloured lines are realisation chosen to span the distribution, and gray lines refer to the deterministic solution.

Even though the probability of remaining in the healthy regime is large, there is still a risk of shifting towards the unhealthy regime despite large reductions in loading. Clearly, these stem from the stochastic perturbations of the model, and the inclusion of noise is therefore important from a management perspective.

The noise is independent of the state and added as Gaussian noise. Although Gaussian noise is often a good first approximation, it fails to obey the basic restrictions of the deterministic model. The problem is that while both states of the deterministic skeleton of the model are positive given that $\boldsymbol{X}_0$ is positive and the loading ($N_{ex,t}$) is greater than zero at all times, the Gaussian noise can produce unrealistic negative results, because of a positive probability of $X_{i,t}+\epsilon_{i,t}$ becoming negative. Evidently, this is likely to happen for the benthic vegetation, when the deterministic solution is close to zero. From a modelling perspective this is clearly not desirable and the next section deals with the inclusion of multiplicative noise to ensure that the states are positive at all times.

2.4 From additive to multiplicative noise

The general situation discussed so far is that the one-step transition of the state distribution is determined by a sum of two Gaussian distributed random variables (the input and the additive noise term), which imply that the one-step distribution is a Gaussian random variable. A more general multivariate formulation of the stochastic model is

$$\boldsymbol{X}_t = \boldsymbol{A}(\boldsymbol{u}_t + \boldsymbol{\epsilon}_t^{\boldsymbol{u}}(\boldsymbol{u}_t)) + \boldsymbol{B}(\boldsymbol{X}_{t-1}) + \boldsymbol{\epsilon}_t^{\boldsymbol{X}}(\boldsymbol{X}_t). \tag{2.10}$$

If $\boldsymbol{\epsilon}_t^{\boldsymbol{u}}(\boldsymbol{u}_t)$ and $\boldsymbol{\epsilon}_t^{\boldsymbol{X}}(\boldsymbol{X}_t)$ are both Gaussian random variables independent of the input and the state, respectively, it is not possible to discriminate between these two contributions, should these both be unknown. To have a biologically meaningful interpretation of the model, we would need at least the following properties of the noise propagation

$$P(\boldsymbol{B}(\boldsymbol{X}_{t-1}) + \boldsymbol{\epsilon}_t^{\boldsymbol{X}}(\boldsymbol{X}_t) < 0) = 0 \tag{2.11}$$

$$P(\boldsymbol{X}_{t|t-1} < 0) = 0. \tag{2.12}$$

Clearly there are many formulations which fulfil these requirements. A simple choice is the log-normal distribution applied in a multiplicative setting

$$\boldsymbol{X}_t = \boldsymbol{A}\boldsymbol{U}_t + \boldsymbol{\Xi}_t\boldsymbol{B}(\boldsymbol{X}_{t-1}), \tag{2.13}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrix functions of appropriate dimensions, $\boldsymbol{U}$ is a log-normal distributed random input, and $\boldsymbol{\Xi}_t$ is a diagonal matrix with diagonal elements $\boldsymbol{\xi}_t$ that follow a multivariate log-normal distribution (Kotz et al., 2000)

Table 2.1: Parameters of the model

Parameter	a_p	$b_{p,b}$	b_b	K	d	t_p	r_p	τ_p	r_b	τ_b
Value	0.8	0.1	0.8	0.2	0.4	0.2	0.4	0.1	0.4	0.1

with expectation equal **1**. Even though many other distributions could have been considered in this setting, the general applicability of the log-normal distribution is well documented (Limpert et al., 2001). Also the choice of the log-normal distribution ensures that first and second moments can always be calculated for finite time horizons, given that the matrix functions $\boldsymbol{A}$ and $\boldsymbol{B}$ are reasonably well behaved, e.g. $\boldsymbol{A} > 0$ and $\boldsymbol{B}$ a linear mass balance function. See also Appendix B of Paper A for a discussion of a choice of distribution that do not have a second order moment representation.

The model presented in Section 2.2 is now formulated within the setting (2.13)

$$\begin{aligned}\begin{bmatrix} X_{p,t} \\ X_{b,t} \end{bmatrix} =& N_{ex,t} \begin{bmatrix} a_p \\ 0 \end{bmatrix} + \begin{bmatrix} \xi_{p,t} & 0 \\ 0 & \xi_{b,t} \end{bmatrix} \times \\ & \left(\begin{bmatrix} b_{p,t-1} & 0 \\ b_{p,b} f_{t-1} & b_b \end{bmatrix} \begin{bmatrix} X_{p,t-1} \\ X_{b,t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ K f_{t-1} \end{bmatrix} \right), \end{aligned} \tag{2.14}$$

where $N_{ex,t}$ is given by

$$N_{ex,t} = \hat{N}_{ex,t} \cdot u_t, \tag{2.15}$$

u_t is a sequence of white noise log-normal distributed random variables, and $\hat{N}_{ex,t}$ is a deterministic loading sequence. Paper A presents a thorough analysis of this model and we will only present some of the main results in the following.

The deterministic parameter values chosen for this analysis are presented in Table 2.1 and simulations are shown in Figure 2.4. The increasing loss of the benthic vegetation with time is evident, and furthermore it should be noted that negative states are not present.

The impact of stochastic perturbations is evident from Figure 2.4. While the deterministic skeleton is in the healthy regime, the stochastic model drifts towards a regime with low benthic biomass and high levels of pelagic biomass. Even though this feature of the model is clear from Figure 2.4 it is useful to derive a more formal definition of the two regimes. To do this let $\boldsymbol{X}_{s,I}$ and $\boldsymbol{X}_{s,II}$ be the two stable equilibrium points of the deterministic model and let $s(X_b)$ be the separator function:

$$\begin{aligned} s(x_b) = \{ & x_p \in \mathbb{R}_0 : \{\boldsymbol{X}_t^{(x_p-\epsilon,x_b)}\} \to X_{s,I}, \{\boldsymbol{X}_t^{(x_p+\epsilon,x_b)}\} \to X_{s,II}, \\ & t \to \infty, \forall \epsilon > 0 \}, \end{aligned} \tag{2.16}$$

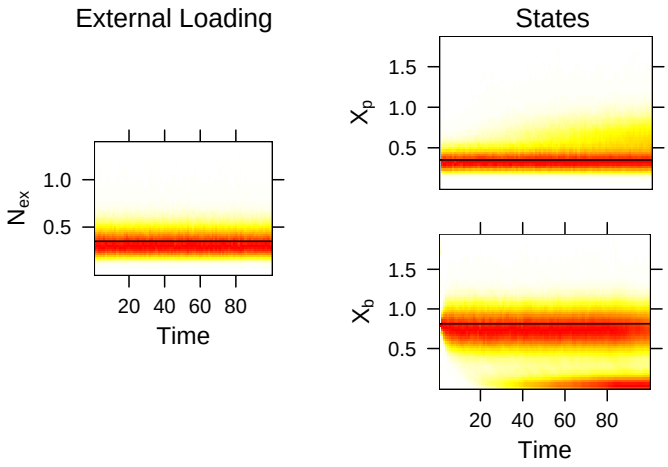

Figure 2.4: System with input noise and system noise. The solid lines are the deterministic response. Noise is simulated by a log-normal distribution with $V[u_t] = 0.1$, $V[\xi_p] = 0.05$, $V[\xi_b] = 0.01$ and $\hat{N}_{ex,t} = 0.35$ for all t. The density plots are based on 5000 realizations of the system and estimated with a Gaussian kernel (the default in "R").

where $\{\boldsymbol{X}_t^{\boldsymbol{y}}\}$ is the deterministic process starting at the point $\boldsymbol{y}$. The healthy and unhealthy regimes are now defined as

$$R_1 = \{(x_b, x_p) \in \mathbb{R}_0^2 : s(x_b) > x_p\} \tag{2.17}$$

$$R_2 = \{(x_b, x_p) \in \mathbb{R}_0^2 : s(x_b) \leq x_p\}, \tag{2.18}$$

where "R_1" is the healthy regime and "R_2" is the unhealthy regime. In general, the separator function needs to be identified numerically, and crossing probabilities need to be identified by simulation studies.

The separator function is shown in Figure 2.5 for different mean values of the loading sequence. For the simplest stochastic model (only input noise), the one-step transition probabilities can be derived analytically (Paper A). These one-step transition probabilities are also shown in Figure 2.5. Even though the one step transition probabilities give an indication of how close the system is to the point of collapse the long term behaviour of the system cannot be derived from plots like Figure 2.5. In addition, the assumption that the system is only subject to random forcing in the input is rather strong, since other random perturbations not accounted for in the model formulation may perturb the states directly (e.g. changing wind conditions). This kind of unaccounted random perturbations should be included as random forcing within the system (by $\boldsymbol{\Xi}_t$).

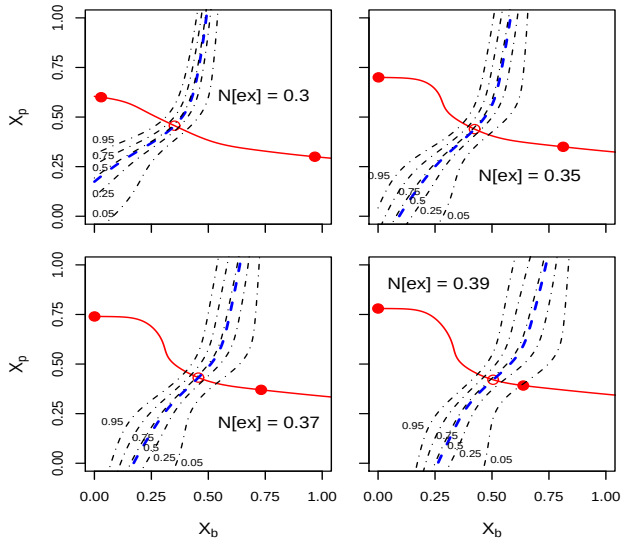

Figure 2.5: Probability of being in regime II in the next time step as a function of the state variables with four different loadings subject to input noise ($V[u_t] = 0.1$ and $u_t \sim LN$), no system noise.

The different impact of noise at different levels of the model is illustrated in Figure 2.6. While input noise and noise in the pelagic level have similar effects on the system, noise in the benthic level have a qualitatively different behaviour and accelerate the transition to the unhealthy regime.

2.5 Discussion

The presented models introduce the concept of regime shifts in discrete time models. The regime shifts are introduced in a setting where we do not argue about the parametric formulation, but introduce functions with the desired properties (the Gaussian CDF). Better parametric descriptions of the regime shifts should be formulated in continuous time, because this is where the dynamics of ecosystem models are formulated. It is, therefore, reasonable not to emphasise the specific parametric formulation of the regime shift function in discrete time.

Considerable emphasis have been put on the noise term and the analysis illus-

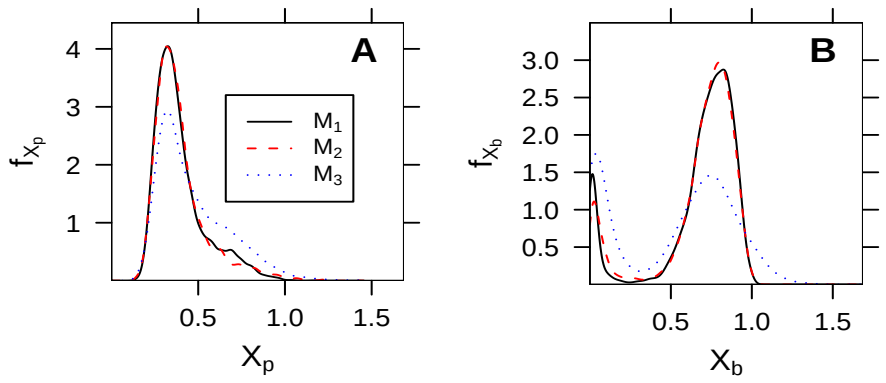

Figure 2.6: Estimated conditional densities after 100 years. All processes start at the point $\boldsymbol{X}_{S,I}$ with a loading of $N_{ex,t} = 0.35$. M_1 is a model with only noise in the input ($V[u_t] = 0.1$), M_2 is the density with noise in the input and in the pelagic level ($V[\xi_p] = 0.05$). Finally, M_3 is the density for a model with noise in the input, the pelagic level, and in the benthic level ($V[\xi_b] = 0.01$).

trate that random perturbations are important for the dynamics of the system. In particular, random perturbations should not change the state-space of the ecosystem. In the examples given here, we introduced multiplicative noise to ensure that the state-space of the stochastic model is the positive real axis.

While the conceptual analysis in this chapter gives suggestions to how regime shifts can be formulated in a mathematical framework, more quantitative methods are needed to verify the existence of regime shift in natural ecosystems. Chapter 3 presents a discrete-time likelihood-based framework for estimation in discrete time models with multiplicative log-normal random noise.

Chapter 3

Discrete time estimation

In general, we cannot observe system states directly and have to rely on noisy observations of the true system. In some cases, all states of the system are observed, but this is often not the case. The opposite might also be true, namely that we have multiple observations of some of the states, e.g. data from multiple monitoring programs. Measurement noise or more generally observation noise originates from the inability to measure the exact state of the system. While system noise as discussed in Chapter 2, are perturbations within the system that affect future states, observation noise does not propagate to future states of the system. The situation is illustrated in Figure 3.1. X_t is the true state of the system and Y_t is the observations of the system. The system is affected by the system noise ϵ_t and an input u_t, while the observations are a function of the current state and the observation error e_t.

In general the stochastic state-space formulation of the situation in Figure 3.1, can be written as

$$\boldsymbol{X}_t = \boldsymbol{f}_{t-1}(\boldsymbol{X}_{t-1}, \boldsymbol{u}_t, \boldsymbol{\epsilon}_t; \boldsymbol{\theta}) \tag{3.1}$$

$$\boldsymbol{Y}_t = \boldsymbol{h}_t(\boldsymbol{X}_t, \boldsymbol{u}_t, \boldsymbol{e}_t; \boldsymbol{\theta}), \tag{3.2}$$

where the state-space of $\boldsymbol{X}_t$ is a subset of $\mathbb{R}^n$, $\boldsymbol{Y}_t$ is a subset of $\mathbb{R}^l$, and $\boldsymbol{f}(\cdot)$ and $\boldsymbol{h}(\cdot)$ are vector functions of appropriate dimensions. $\boldsymbol{\epsilon}_t$ and $\boldsymbol{e}_t$ are random variables and $\boldsymbol{\theta}$ is a parameter vector to be estimated.

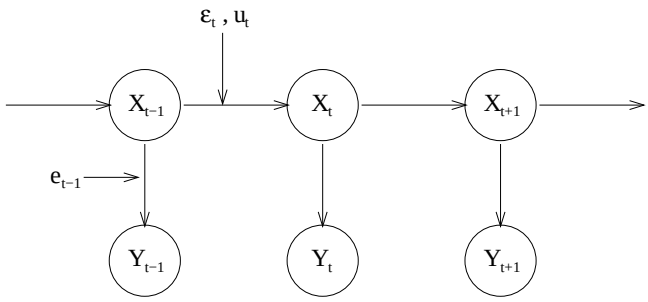

Figure 3.1: Schematic illustration of the stochastic state-space formulation. X_t is a hidden state which is observed under observation noise e_t, while the state progression is affected by system noise (ϵ_t).

The estimation problem can be divided into two parts; 1) the filtering problem, where the state of the true system $\hat{\boldsymbol{X}}_{t|t-1}$ is estimated given the information provided by the observations $\mathcal{Y}_{t-1} = \{\boldsymbol{Y}_1, ..., \boldsymbol{Y}_{t-1}\}$ up to time $t-1$, and the parameters $\boldsymbol{\theta}$, and 2) the estimation problem, where the best estimate, $\hat{\boldsymbol{\theta}}$, of $\boldsymbol{\theta}$ based on the observations $\mathcal{Y}_N$ is inferred. In the present context, the meaning of the term "best" is that $\hat{\boldsymbol{\theta}}$ maximises the likelihood of $\mathcal{Y}_N$.

Since the process (3.1) is a Markov process, the one-step prediction density can be written as

$$p(\boldsymbol{X}_t|\mathcal{Y}_{t-1}) = p(\boldsymbol{X}_t|\boldsymbol{Y}_{t-1})p(\boldsymbol{X}_{t-1}|\mathcal{Y}_{t-2}) \tag{3.3}$$

$$\vdots$$

$$= p(\boldsymbol{X}_0)\prod_{i=1}^{t} p(\boldsymbol{X}_i|\boldsymbol{Y}_{i-1}), \tag{3.4}$$

with each of the conditional densities given by

$$p(\boldsymbol{X}_i|\boldsymbol{Y}_{i-1}) = \int p(\boldsymbol{X}_i|\boldsymbol{X}_{i-1})p(\boldsymbol{X}_{i-1}|\boldsymbol{Y}_{i-1})d\boldsymbol{X}_{i-1}. \tag{3.5}$$

The conditional density $p(\boldsymbol{X}_i|\boldsymbol{Y}_i)$ is referred to as the reconstruction and is obtained from Bayes rule

$$p(\boldsymbol{X}_i|\boldsymbol{Y}_i) = \frac{p(\boldsymbol{Y}_i|\boldsymbol{X}_i)p(\boldsymbol{X}_i|\boldsymbol{Y}_{i-1})}{p(\boldsymbol{Y}_i|\boldsymbol{Y}_{i-1})}, \tag{3.6}$$

and finally the denominator is calculated by the integral

$$p(\boldsymbol{Y}_i|\boldsymbol{Y}_{i-1}) = \int p(\boldsymbol{Y}_i|\boldsymbol{X}_i)p(\boldsymbol{X}_i|\boldsymbol{Y}_{i-1})d\boldsymbol{X}_i. \tag{3.7}$$

The equations (3.5)-(3.7) constitute a recursive Bayesian filter. For general noise processes and nonlinear functions $\boldsymbol{f}(\cdot)$ and $\boldsymbol{h}(\cdot)$ the solution needs to be obtained by simulation based methods like MCMC-methods or particle filters (e.g. Givon et al. (2009)).

3.1 Kalman filtering and likelihood estimation

If $\boldsymbol{f}(\cdot)$ and $\boldsymbol{h}(\cdot)$ are linear in $\boldsymbol{X}_t$, and $\boldsymbol{e}_t$ and $\boldsymbol{\epsilon}_t$ are Gaussian white noise sequences, the Bayesian filter amounts to the classical Kalman filter (Kalman, 1960). While for $\boldsymbol{f}$ and $\boldsymbol{h}$ being nonlinear functions, and $\boldsymbol{e}_t$ and $\boldsymbol{\epsilon}_t$ Gaussian white noise sequences, approximate filters like the Extended Kalman Filter (EKF) (Jazwinski (1970), in an additive noise setting) are available. For more general noise structures the EKF is a second order moment approximation, but the distributional properties need not be well approximated by Gaussian densities.

The following discrete-time EKF (due to Simon (2006)), including general noise dependence of the state and observation equation, is given for completeness.

Theorem 3.1 Discrete-time Extended Kalman Filter: *Given the system and observation equation*

$$\boldsymbol{X}_t = \boldsymbol{f}_{t-1}(\boldsymbol{X}_{t-1}, \boldsymbol{u}_{t-1}, \boldsymbol{\epsilon}_t; \boldsymbol{\theta}) \tag{3.8}$$
$$\boldsymbol{Y}_t = \boldsymbol{h}_t(\boldsymbol{X}_t, \boldsymbol{u}_t, \boldsymbol{e}_t; \boldsymbol{\theta}) \tag{3.9}$$
$$\boldsymbol{\epsilon}_t \sim D(\boldsymbol{0}, \boldsymbol{\Sigma}_t^{\epsilon}) \tag{3.10}$$
$$\boldsymbol{e}_t \sim D(\boldsymbol{0}, \boldsymbol{S}_t), \tag{3.11}$$

where $\sim D(\boldsymbol{0}, \boldsymbol{\Sigma})$ *refers to a random variable with a second-order moment representation, expectation* $\boldsymbol{0}$*, covariance matrix* $\boldsymbol{\Sigma}$*, and the initial condition*

$$\boldsymbol{X}_{1|0} = \boldsymbol{\mu}_0^{\boldsymbol{X}}, \quad \hat{\boldsymbol{\Sigma}}_{1|0}^{xx} = \boldsymbol{V}_0. \tag{3.12}$$

Then the state and state covariance reconstruction is given by

$$\hat{\boldsymbol{X}}_{t|t} = \hat{\boldsymbol{X}}_{t|t-1} + \boldsymbol{K}_t(\boldsymbol{Y}_t - \boldsymbol{h}_t(\hat{\boldsymbol{X}}_{t|t-1}, \boldsymbol{u}_t, \boldsymbol{0}; \boldsymbol{\theta})) \tag{3.13}$$
$$\boldsymbol{\Sigma}_{t|t}^{xx} = (\boldsymbol{I} - \boldsymbol{K}_t\boldsymbol{H}_t)\boldsymbol{\Sigma}_{t|t-1}^{xx} \tag{3.14}$$

and the state prediction and covariance matrix are given by

$$\hat{\boldsymbol{X}}_{t+1|t} = \boldsymbol{f}_t(\boldsymbol{X}_{t|t}, \boldsymbol{u}_t, \boldsymbol{0}; \boldsymbol{\theta}) \tag{3.15}$$
$$\boldsymbol{\Sigma}_{t+1|t}^{xx} = \boldsymbol{F}_t\boldsymbol{\Sigma}_{t|t}^{xx}\boldsymbol{F}_t^T + \boldsymbol{L}_t\boldsymbol{\Sigma}_t^{\epsilon}\boldsymbol{L}_t^T \tag{3.16}$$
$$\boldsymbol{\Sigma}_{t+1|t}^{yy} = \boldsymbol{H}_t\boldsymbol{\Sigma}_{t|t-1}^{xx}\boldsymbol{H}_t^T + \boldsymbol{M}_t\boldsymbol{S}_t\boldsymbol{M}_t^T, \tag{3.17}$$

where the Kalman gain K_t is given by

$$\boldsymbol{K}_t = \boldsymbol{\Sigma}^{xx}_{t|t-1}\boldsymbol{H}_t^T \left(\boldsymbol{\Sigma}^{yy}_{t|t-1}\right)^{-1}, \tag{3.18}$$

and $\boldsymbol{F}_t$, $\boldsymbol{L}_t$, $\boldsymbol{H}_t$ and $\boldsymbol{M}_t$ given by the Jacobians

$$\begin{aligned} \boldsymbol{F}_t &= \left.\frac{\partial \boldsymbol{f}_t}{\partial \boldsymbol{x}}\right|_{\boldsymbol{X}_{t|t}}, \qquad \boldsymbol{L}_t = \left.\frac{\partial \boldsymbol{f}_t}{\partial \boldsymbol{\epsilon}}\right|_{\boldsymbol{X}_{t|t}}, \\ \boldsymbol{H}_t &= \left.\frac{\partial \boldsymbol{h}_t}{\partial \boldsymbol{x}}\right|_{\boldsymbol{X}_{t|t-1}}, \qquad \boldsymbol{M}_t = \left.\frac{\partial \boldsymbol{h}_t}{\partial \boldsymbol{e}}\right|_{\boldsymbol{X}_{t|t-1}}. \end{aligned} \tag{3.19}$$

The EKF is a linear approximation of the true nonlinear dynamics of the system and generalise the classical linear Kalman filter. If $\boldsymbol{f}_t$ and $\boldsymbol{h}_t$ are both linear in $\boldsymbol{X}_t$, and the system and observation noises are both additive and Gaussian, the Kalman filter is exact. In this case the likelihood function is given by (Madsen, 2008)

$$L(\boldsymbol{\theta}; \mathcal{Y}_t) = \prod_{t=1}^{N} \frac{1}{\sqrt{(2\pi)^m \det(\boldsymbol{\Sigma}^{yy}_{t|t-1})}} \exp\left(-\frac{1}{2}\tilde{\boldsymbol{e}}_t^T \left(\boldsymbol{\Sigma}^{yy}_{t|t-1}\right)^{-1} \tilde{\boldsymbol{e}}_t\right), \tag{3.20}$$

where the innovation error ($\tilde{\boldsymbol{e}}_t$) is given by

$$\tilde{\boldsymbol{e}}_t = \boldsymbol{Y}_t - \hat{\boldsymbol{Y}}_{t|t-1}. \tag{3.21}$$

If the state and observation equations (3.8)-(3.9) can be assumed to be approximately local Gaussian, then (3.20) can form the basis for approximate maximum likelihood estimation. This is the key assumption in e.g. Kristensen et al. (2004a). Furthermore the assumption in this reference is further that the system noise is independent of the state, i.e. $\boldsymbol{f}_{t-1}$ splits into a sum of two functions, one that is independent of $\boldsymbol{\epsilon}_t$ and one independent of X_{t-1}.

3.2 Estimation with multiplicative noise

As pointed out in Chapter 2, the log-normal distribution plays an important role in many applications and the log-normal assumption is standard in many biological ecological applications. The models derived in Section 2.4 use this assumption together with additive inputs, which is essential for mass balance models. Even though the mass balance arguments are important for the deterministic formulation of the problem, these cannot be expected to hold for the stochastic perturbations in an open system.

In this section the results from Paper B are summarised while the detailed derivation are omitted. Assuming that the observations are log-normal distributed around the true state of the system, the state-space formulation of the linear version of the models presented in Section 2.4 is given by

$$\boldsymbol{X}_t = \boldsymbol{\Xi}_t \boldsymbol{A}_t \boldsymbol{X}_{t-1} + \boldsymbol{B}_t \boldsymbol{U}_{t-1}, \tag{3.22}$$

$$\boldsymbol{Y}_t = \boldsymbol{\Lambda}_t \boldsymbol{C}_t \boldsymbol{X}_t + \boldsymbol{\epsilon}_t, \tag{3.23}$$

where the matrices $\boldsymbol{A}_t$, $\boldsymbol{B}_t$ and $\boldsymbol{C}_t$ are allowed to depend on time, $\boldsymbol{U}_t$ is a random vector with mean $\boldsymbol{\mu}_t^{\boldsymbol{u}}$ and covariance matrix $\boldsymbol{\Sigma}_t^{\boldsymbol{u}}$ (allowed to be zero), $\boldsymbol{\Lambda}_t$ and $\boldsymbol{\Xi}_t$ are diagonal matrices with diagonal elements containing the random vectors $\boldsymbol{\lambda}_t$ and $\boldsymbol{\xi}_t$ with expectation equal $\mathbf{1}$ and covariances $\boldsymbol{\Sigma}_t^{\boldsymbol{\xi}}$ and $\boldsymbol{\Sigma}_t^{\boldsymbol{\lambda}}$, respectively, and $\boldsymbol{\epsilon}_t$ is a random variable with mean $\boldsymbol{\mu}_t^{\boldsymbol{\epsilon}}$ and covariance matrix $\boldsymbol{\Sigma}_t^{\epsilon}$.

The introduction of the additive observation noise term $\boldsymbol{\epsilon}_t$, ensures that the filter reconstruction can be calculated even when the estimation render state estimates equal zero (of special importance when optimising the likelihood). On the other hand it is also reasonable from an observational point of view, in the sense that even when the state is zero, measurement noise will still often be present. The notational shift between $\boldsymbol{A}$ and $\boldsymbol{B}$ (compared with Chapter 2) is due to a change of notation between Paper A and B, the notation in the present chapter coincide with the notation used in Paper B.

3.2.1 State estimation

State-space models with multiplicative noise have been studied earlier in different settings e.g. Zhang & Zhang (2007) (multiplicative observation noise) and Jimenez & Ozaki (2002) (continuous time multiplicative and additive noise). In Paper B the projection theorem (Madsen, 2008) is used to derive an approximate filter for the stochastic state-space model (3.22)-(3.23), and the result is summarised below.

Theorem 3.2 Multiplicative filter: *Given the discrete-time state-space model in Eqs.* (3.22)-(3.23)*, with given initial conditions and given initial conditions*

$$\hat{\boldsymbol{X}}_{1|0} = \boldsymbol{\mu}_0, \quad \boldsymbol{\Sigma}_{1|0}^{xx} = \boldsymbol{V}_0. \tag{3.24}$$

the approximate state reconstruction is given by

$$\hat{\boldsymbol{X}}_{t|t} = \hat{\boldsymbol{X}}_{t|t-1} + \boldsymbol{K}_t(\boldsymbol{Y}_t - \hat{\boldsymbol{Y}}_{t|t-1}) \tag{3.25}$$

$$\boldsymbol{\Sigma}_{t|t}^{xx} = \boldsymbol{\Sigma}_{t|t-1}^{xx} - \boldsymbol{K}_t(\boldsymbol{\Sigma}_{t|t-1}^{xy})^T. \tag{3.26}$$

and the predictions are given by

$$\hat{\boldsymbol{X}}_{t+1|t} = \boldsymbol{A}_{t+1}\hat{\boldsymbol{X}}_{t|t} + \boldsymbol{B}_{t+1}\boldsymbol{\mu}_t^{\boldsymbol{u}} \tag{3.27}$$

$$\boldsymbol{\Sigma}_{t+1|t}^{xx} = \boldsymbol{A}_{t+1}\boldsymbol{\Sigma}_{t|t}^{xx}\boldsymbol{A}_{t+1}^T + \boldsymbol{\Sigma}_{t+1}^{\boldsymbol{\xi}} \odot (\boldsymbol{A}_{t+1}\boldsymbol{P}_{t|t}^{xx}\boldsymbol{A}_{t+1}^T) + \boldsymbol{B}_{t+1}\boldsymbol{\Sigma}_t^{\boldsymbol{u}}\boldsymbol{B}_{t+1}^T \tag{3.28}$$

$$\hat{\boldsymbol{Y}}_{t+1|t} = \boldsymbol{C}_{t+1}\hat{\boldsymbol{X}}_{t+1|t} + \boldsymbol{\mu}_t^{\boldsymbol{\epsilon}} \tag{3.29}$$

$$\boldsymbol{\Sigma}_{t+1|t}^{yy} = \boldsymbol{C}_{t+1}\boldsymbol{\Sigma}_{t+1|t}^{xx}\boldsymbol{C}_{t+1}^T + \boldsymbol{\Sigma}_{t+1}^{\boldsymbol{\lambda}} \odot (\boldsymbol{C}_{t+1}\boldsymbol{P}_{t+1|t}^{xx}\boldsymbol{C}_{t+1}^T) + \boldsymbol{\Sigma}_{t+1}^{\boldsymbol{\epsilon}} \tag{3.30}$$

$$\boldsymbol{\Sigma}_{t+1|t}^{xy} = \boldsymbol{\Sigma}_{t+1|t}^{xx}\boldsymbol{C}_{t+1}^T, \tag{3.31}$$

where $\odot$ *refer to element-wise multiplication and*

$$\boldsymbol{K}_t = \boldsymbol{\Sigma}_{t|t-1}^{xx}\boldsymbol{C}_t^T(\boldsymbol{\Sigma}_{t|t-1}^{yy})^{-1} \tag{3.32}$$

$$\boldsymbol{P}_{t+k|t}^{xx} = \boldsymbol{\Sigma}_{t+k|t}^{xx} + \hat{\boldsymbol{X}}_{t+k|t}\hat{\boldsymbol{X}}_{t+k|t}^T. \tag{3.33}$$

If the dynamics is such that the process is positive for all t *then (3.25) is replaced by*

$$(\hat{\boldsymbol{X}}_{t|t})_i = \max\{0, (\hat{\boldsymbol{X}}_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}^{xy}(\boldsymbol{\Sigma}_{t|t-1}^{yy})^{-1}(\boldsymbol{Y}_t - \hat{\boldsymbol{Y}}_{t|t-1}))_i\}. \tag{3.34}$$

When the state-space of $\boldsymbol{X}_t$ is $\mathbb{R}_0^n$ Eq. (3.34) ensures that the reconstruction stays within the state-space for all t. The filter equations given above are approximations, because the reconstruction is based on the linear approximation in the projection theorem. The simulation examples presented in Paper B and Møller et al. (2009) do, however, yield convincing results.

3.2.2 Likelihood estimation

To be able to construct the likelihood function for a given set of parameters, we need to impose some distributional assumption on the observations of the system given above. An obvious candidate is the log-normal distribution, which is also supported by Fenton (1960), who report that the sum of log-normal distributed random variables can be approximate by another log-normal distributed random variable with mean and variance depending on the sum of the moments of the original distributions. In the present context the parametric formulation of the moments is not important because these are estimated by the filter equations.

Under the assumption that $\hat{\boldsymbol{Y}}_{t+1|t}$ follows a multivariate log-normal distribution, $\hat{\boldsymbol{Z}}_{t+1|t} = \log(\hat{\boldsymbol{Y}}_{t+1|t})$ follows a multivariate normal distribution with mean and

variance given by (Kotz et al. (2000) and Paper B)

$$\boldsymbol{\Sigma}^{zz}_{t+1|t} = \log\left(\boldsymbol{\Sigma}^{yy}_{t+1|t} + \hat{\boldsymbol{Y}}_{t+1|t}\hat{\boldsymbol{Y}}^{T}_{t+1|t}\right) - \log\left(\hat{\boldsymbol{Y}}_{t+1|t}\hat{\boldsymbol{Y}}^{T}_{t+1|t}\right) \tag{3.35}$$

$$\hat{\boldsymbol{Z}}_{t+1|t} = \log\left(\hat{\boldsymbol{Y}}_{t+1|t}\right) - \frac{1}{2}\mathrm{diag}\boldsymbol{\Sigma}^{zz}_{t+1|t}, \tag{3.36}$$

where the logarithm is applied element-wise in (3.35).

The covariance matrices in Theorem 3.2 are calculated using the projection theorem and as such the filter equations guaranties that the covariance matrix are positive definite and that the correlations (ρ^{yy}_{ij}) are given by

$$\rho^{yy}_{ij} = \frac{\left(\boldsymbol{\Sigma}^{yy}_{t+1|t}\right)_{i,j}}{\sqrt{\left(\boldsymbol{\Sigma}^{yy}_{t+1|t}\right)_{i,i}\left(\boldsymbol{\Sigma}^{yy}_{t+1|t}\right)_{j,j}}} \in (-1,1). \tag{3.37}$$

Even though (3.37) is a correlation matrix for a multivariate Gaussian random variable, it might not be a correlation matrix for a multivariate log-normal random variable. A small example can illustrate this point.

Example 3.1 *Suppose we are given a multivariate (supposedly)* log*-normal distributed random variable* $\boldsymbol{Y}$*, with mean and variance equal*

$$E[\boldsymbol{Y}] = \begin{bmatrix} 1 \\ 0.1 \end{bmatrix}, \quad V[\boldsymbol{Y}] = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}, \tag{3.38}$$

$V[\boldsymbol{Y}]$ *is clearly a covariance matrix, but calculating the covariance matrix for the corresponding Gaussian random variable* $\boldsymbol{Z} = \log(\boldsymbol{Y})$ *using* (3.35) *gives*

$$V[\boldsymbol{Z}] = \begin{bmatrix} \log(2.0) & \log(10) \\ \log(10) & \log(101) \end{bmatrix}. \tag{3.39}$$

A calculation of the correlation gives $\rho^{z}_{12} = \frac{\log(10)}{\sqrt{\log(2)\log(101)}} = 1.29$*, which shows that* $V[\boldsymbol{Z}]$ *is not a covariance matrix.* □

Covariance matrices produced by the filter equations in Theorem 3.2, that do not have a covariance matrix for the corresponding Gaussian distribution are referred to as not admissible. Given that the expectation $\hat{\boldsymbol{Y}}_{t|t-1}$ and diagonal elements of the covariance matrix in (3.30) are captured well, it is possible to calculate the interval of admissible correlations ($\rho^{yy}_{ij,min}, \rho^{yy}_{ij,max}$) for the covariance

matrix $\mathbf{\Sigma}^{yy}_{t|t-1}$, and thereby defining a transformation of the cross covariance term based on diagonal elements of (3.35)

$$\tilde{\sigma}^{yy}_{ij}(\rho^{yy}_{ij}) = \hat{Y}_{i,t+1|t}\hat{Y}_{j,t+1|t}\left(e^{\rho^{yy}_{ij}\sqrt{\sigma^{zz}_{ii}\sigma^{zz}_{jj}}} - 1\right). \tag{3.40}$$

The transformed covariance matrix $\left(\tilde{\mathbf{\Sigma}}^{yy}_{t|t-1}\right)$ is now defined as

$$\left(\tilde{\mathbf{\Sigma}}^{yy}_{t|t+1}\right)_{i,j} = \begin{cases} \sigma^{yy}_{ii} & \text{for} \quad i = j \\ \tilde{\sigma}^{yy}_{ij} & \text{for} \quad i \neq j. \end{cases} \tag{3.41}$$

The transformation (3.41) is monotone and the covariance matrix of $\boldsymbol{Z}_{t+1|t}$ is defined as

$$\tilde{\mathbf{\Sigma}}^{zz}_{t+1|t} = \log\left(\tilde{\mathbf{\Sigma}}^{yy}_{t+1|t} + \hat{\boldsymbol{Y}}_{t+1|t}\hat{\boldsymbol{Y}}^{T}_{t+1|t}\right) - \log\left(\hat{\boldsymbol{Y}}_{t+1|t}\hat{\boldsymbol{Y}}^{T}_{t+1|t}\right). \tag{3.42}$$

The transformation guaranties that $\tilde{\mathbf{\Sigma}}^{zz}_{t+1|t}$ is a covariance matrix. Further $\rho^{yy}_{ij} = \rho^{yy}_{ij,min} \Rightarrow \rho^{zz}_{ij} = -1$, $\rho^{yy}_{ij} = 0 \Rightarrow \rho^{zz}_{ij} = 0$ or $\rho^{yy}_{ij} = \rho^{yy}_{ij,max} \Rightarrow \rho^{zz}_{ij} = 1$. Defining the innovation $\tilde{\boldsymbol{Z}}_{t+1|t}$ as

$$\tilde{\boldsymbol{Z}}_{t+1|t} = \log\left(\boldsymbol{Y}_{t+1}\right) - \log\left(\hat{\boldsymbol{Y}}_{t+1|t}\right) + \frac{1}{2}\mathrm{diag}\tilde{\mathbf{\Sigma}}^{zz}_{t+1|t}, \tag{3.43}$$

the log-likelihood becomes

$$\log L(\boldsymbol{\theta}; \mathcal{Y}_N) = -\frac{1}{2}\sum_{i=1}^{N}\left(\log(\det\tilde{\mathbf{\Sigma}}^{zz}_{i|i-1}) + \tilde{\boldsymbol{Z}}^{T}_{i|i-1}(\tilde{\mathbf{\Sigma}}^{zz}_{i|i-1})^{-1}\tilde{\boldsymbol{Z}}_{i|i-1}\right) + c, \tag{3.44}$$

where c is a constant independent of $\mathcal{Y}_t$ and $\boldsymbol{\theta}$. The maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \arg\left\{\max_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta}; \mathcal{Y}_N)\right\}. \tag{3.45}$$

Paper B presents simulation based examples of the presented procedure based on the log-normal assumption and a comparison with a local Gaussian assumption. Further examples are presented in Møller et al. (2009), where the log-normal assumption is not valid (only state estimation) and further comments on state estimation with missing or partially missing data are given in that reference.

3.3 Discussion

The state and parameter estimation problem is important in many applications; also in the context of marine ecosystem. The filter equations derived in this chapter provide the basis of maximum likelihood estimation in systems with multiplicative log-normal noise and additive inputs. The additive nature of the input is important for mass balance arguments in the deterministic skeleton of the model and therefore in dynamical models for ecosystems. The additive input does however make direct transformation (in this case the log-transform) of the state-space complicated to apply directly in the EKF setting. To be able to infer reliable parameter estimation, we need to assume some properties of the transition probabilities and a key result is the direct formulation of the likelihood function.

The filter equations are derived in a linear dynamical formulation and as showed in Chapter 2 nonlinear dynamical equations are needed to formulate hypothesises of ecological regime shifts. Given the formulation of the filter equations in this chapter the transition to nonlinear models as presented in Chapter 2 is in principle not very complicated, since the dynamic matrix $\boldsymbol{A}(\boldsymbol{X}_t)$ can be replaced by the corresponding Jacobian to obtain an EKF-like set up for the estimation procedure.

The model formulation and estimation so far, has focused on discrete time modelling under the hypothesis of log-normal transition probabilities. The transition to continuous time models paves the way for a much richer transition probability structure. The log-normal assumption used in this chapter implies that the variance scales with the squared state, corresponding to a random environment (Tier & Hanson, 1981). As we will see in Chapter 4 local variance scaling with the squared state might be too strong and choosing processes such that the variance scales linearly with the state, corresponding to demographic noise (Tier & Hanson, 1981), might be a better approach.

Chapter 4

Modelling by Stochastic Differential Equations

From a modelling point of view, the transition from discrete-time to continuous-time offers a more simple conceptual setup, because arguments of mass flows between states apply directly in the differential setting. In discrete time, biomass or nutrients can move through several states/trophic levels within one time step, while in continuous time biomass can only move up one trophic level. A consequence of this is that discrete-time models might need a considerably larger number of parameters than the continuous-time equivalent model (see Example 4.1 below). The continuous-time formulation of random noise does, however, require more attention, because independence of the noise process does not lead to reasonable interpretation of the resulting differential equation (Øksendal, 2003).

Assuming that we are given an ordinary differential equation (ODE) and want to formulate this in a stochastic setting. The first formulation could be

$$\frac{d\boldsymbol{x}_t}{dt} = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta}) + \boldsymbol{w}_t, \tag{4.1}$$

where the first part is the usual ODE part and $\boldsymbol{w}_t$ is a stochastic perturbation. As the first approach, it is natural to impose similar requirements on $\boldsymbol{w}_t$ as we do in a discrete time setting, i.e. $E[\boldsymbol{w}_t] = 0$, $\forall t$, and $\boldsymbol{w}_{t_1}$ and $\boldsymbol{w}_{t_2}$ independent

if $t_1 \neq t_2$. It does, however, turn out that the second requirement leads to a process with either infinite or zero variance (Øksendal, 2003).

A more reasonable requirement is therefore $\boldsymbol{w}_{t_2} - \boldsymbol{w}_{t_1}$ independent of $\boldsymbol{w}_{t_4} - \boldsymbol{w}_{t_3}$ when $[t_1, t_2] \cap [t_3, t_4] = \emptyset$. This implies that the process $\boldsymbol{w}_t$ has independent increments rather than $\boldsymbol{w}_t$ being a white noise process itself. For general processes $\boldsymbol{w}_t$ the integral formulation (and interpretation) of (4.1) defines the class of Lévy processes, and when $\boldsymbol{w}_t$ is Gaussian, this is referred to as a Stochastic Differential equation (SDE), and the usual notation is (Øksendal, 2003)

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t; \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})d\boldsymbol{w}_t, \tag{4.2}$$

where $t \in \mathbb{R}_0$ is time, $\boldsymbol{x}_t \in \mathbb{R}^n$ is the state, $\boldsymbol{u}_t \in \mathbb{R}^r$ is an input vector, $\boldsymbol{f}(\cdot)$ is a vector function into $\mathbb{R}^n$, $\boldsymbol{\sigma}(\cdot)$ is a matrix function into $\mathbb{R}^{n \times m}$, $\boldsymbol{w}_t \in \mathbb{R}^m$ is the standard Brownian motion (i.e. $\boldsymbol{w}_t \sim N(0, \boldsymbol{I}t)$), and $\boldsymbol{\theta} \in \mathbb{R}^p$ is a parameter vector.

Since the continuous-time Brownian motion is nowhere differentiable (Karatzas & Shreve, 1991) the formulation (4.2) has no meaning without an integral interpretation. There are two convenient choices of integral interpretation; the Itô interpretation and the Stratonovich interpretation. In general, these do not coincide. While the Stratonovich interpretation leads to the ordinary chain rule under transformation, the Itô interpretation does not. On the other hand the Itô interpretation does not use information on future (unknown) values of the Brownian motion in the integration (Øksendal, 2003). Further Stratonovich integrals are not Martingales, while Itô integrals are. Øksendal (2003) points out that the point on future information makes the Itô interpretation more reasonable in many applications. The martingale property is important in filtering theory. We will therefore restrict the analysis to the Itô integral interpretation.

Just as in the discrete-time case, one cannot observe $\boldsymbol{x}_t$ directly, but will have to rely on noise measurements of the true system. In the discrete-time case we had an underlying discrete-time system and a set of discrete-time observations. In this chapter, the underlying process is evolving in continuous time, and observations are functions of the true state sampled in discrete time. In general, the type of problem we are faced with is: Given a set of discrete-time observations $\mathcal{Y}_N = \{\boldsymbol{y}_1, ..., \boldsymbol{y}_N\}$ find the best set of parameters $\boldsymbol{\theta}$ *and* the best structure of $\boldsymbol{f}(\cdot)$, $\boldsymbol{h}(\cdot)$ and $\boldsymbol{\sigma}(\cdot)$. This problem can be formalised by the continuous-discrete time stochastic state-space formulation (Jazwinski (1970))

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t; \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})d\boldsymbol{w}_t \tag{4.3}$$

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{u}_k, t_k, \boldsymbol{e}_k; \boldsymbol{\theta}), \tag{4.4}$$

where $\boldsymbol{y}_k$ is the observation at time t_k, $\boldsymbol{u}_k$ is the input at time t_k, $\{\boldsymbol{e}_k\}_{k=1}^N$ is a sequence of random variables belonging to the distribution $D_k(\boldsymbol{\theta})$. Paper C

deals with estimation in a simple model when the structure of $\boldsymbol{f}(\cdot)$, $\boldsymbol{\sigma}(\cdot)$ and $\boldsymbol{h}(\cdot)$ is assumed known, Paper F deals with estimation of $\boldsymbol{\theta}$ and identification of the structure of $\boldsymbol{f}(\cdot)$ and $\boldsymbol{\sigma}(\cdot)$, while Paper G includes identification of the structure of $\boldsymbol{h}(\cdot)$.

Before we turn to the solution of SDEs, the next example illustrates how the continuous-time formulation can lead to fewer parameters when the dimension of the state-space is high.

Example 4.1 *Paper F presents the development of a nonlinear SDE model for phytoplankton growth. The nitrogen pool is divided into two parts: 1) phytoplankton nitrogen and 2) nitrogen not included in phytoplankton. With the simple two state model there is no gain in terms of the number of parameters to be estimated when modelling in continuous time. In the discussion of Paper F it is however suggested to further split the nitrogen pool (not contained in phytoplankton) into two parts (organic and inorganic nitrogen). The conceptual setting is illustrated in Figure 4.1, and the mathematical formulation of the linearised dynamics is*

$$\begin{aligned}\begin{bmatrix} X_{o,t} \\ X_{i,t} \\ X_{p,t} \end{bmatrix} &= \begin{bmatrix} -a_{oi} - Q_t - a_{wl} & 0 & a_{po} \\ a_{oi} & -a_{ip} - Q_t & 0 \\ 0 & a_{ip} & -a_{pi} - Q_t \end{bmatrix} \begin{bmatrix} X_{o,t} \\ X_{i,t} \\ X_{p,t} \end{bmatrix} dt + \\ &\quad \begin{bmatrix} 0 \\ N_{ex,t} Q_t \\ 0 \end{bmatrix} dt + \boldsymbol{\sigma}(\boldsymbol{X}_t) d\boldsymbol{w}_t \qquad (4.5) \\ &= (\boldsymbol{A}\boldsymbol{X}_t + \boldsymbol{B} N_{ex,t} Q_t)\, dt + \boldsymbol{\sigma}(\boldsymbol{X}_t) d\boldsymbol{w}_t \qquad (4.6)\end{aligned}$$

The drift term of the linear model has 5 parameters (the elements of $\boldsymbol{A}$) to be estimated, if we were to solve this in a discrete time setting then the transition matrix involves the matrix exponential of $\boldsymbol{A}$, which is a dense matrix and the number of parameters is equal the number of elements in $\boldsymbol{A}$ namely 9. In addition, the loading is assumed to be purely inorganic nitrogen, and in the discrete time setting we would need to describe how this is distributed among the states and how much is lost due to outflow and sediment burial resulting in an additional 3 parameters. □

Example 4.1 illustrates one of the key advantages of the continuous-time modelling and why it is worthwhile to consider the continuous-time counterpart of discrete time stochastic models. Another important issue is the noise propagation. Local Gaussian random increments give rise to a very rich class of transition probability distributions that might be very difficult to infer in discrete time.

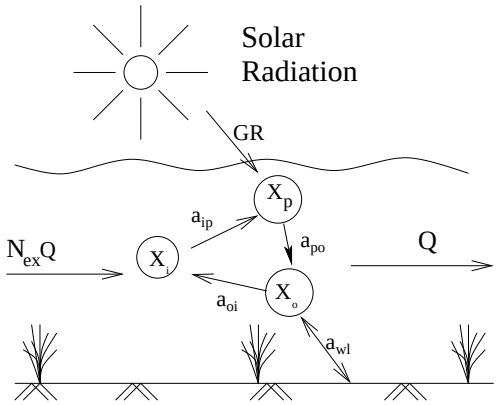

Figure 4.1: Conceptual setting for Example 4.1. The assumptions behind the model is: the water column is enriched by inorganic nitrogen ($N_{ex,t}$) due to freshwater runoff (Q_t), phytoplankton (X_p) growth is based on inorganic nitrogen (X_i), dead phytoplankton contributes to the organic nitrogen pool (X_o), and organic nitrogen is converted to inorganic nitrogen by bacterial processes. Further all processes are subject to losses through outflow $Q_t X_t$ and organic nitrogen is lost from sedimentation.

4.1 Solutions of Stochastic Differential Equations

For simple SDEs, solutions of (4.2) can be calculated by change of variable (Itô's lemma), change of measure (Girsanov theorem) (Øksendal, 2003) or random time change (Klebaner, 2005). More generally the transition probabilities for the stochastic process $\boldsymbol{x}_t$ are given by the Fokker-Planck (Kolmogorov forward) equation (Gard, 1988)

$$\begin{aligned}\frac{\partial p(\boldsymbol{x},t|\boldsymbol{x}_0)}{\partial t} = &-\sum_{i=1}^{n}\frac{\partial}{\partial x_i}f_i(\boldsymbol{x},t)p(\boldsymbol{x},t|\boldsymbol{x}_0)+\\ &\frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i\partial x_j}(\boldsymbol{\sigma}(\boldsymbol{x},t)\boldsymbol{\sigma}(\boldsymbol{x},t)^T)_{ij}p(\boldsymbol{x},t|\boldsymbol{x}_0),\end{aligned} \tag{4.7}$$

or in a filtering framework

$$\begin{aligned}\frac{\partial p(\boldsymbol{x},t|\mathcal{Y}_{t_k})}{\partial t} = &-\sum_{i=1}^{n}\frac{\partial}{\partial x_i}f_i(\boldsymbol{x},t)p(\boldsymbol{x},t|\mathcal{Y}_{t_k})+\\ &\frac{1}{2}\sum_{i,j}\frac{\partial^2}{\partial x_i\partial x_j}(\boldsymbol{\sigma}(\boldsymbol{x},t)\boldsymbol{\sigma}(\boldsymbol{x},t)^T)_{ij}p(\boldsymbol{x},t|\mathcal{Y}_{t_k}),\end{aligned} \tag{4.8}$$

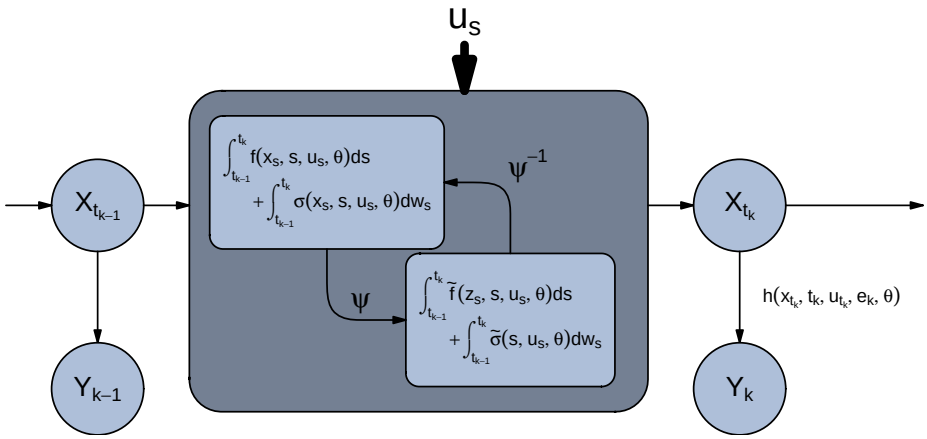

Figure 4.2: Conceptual diagram of the estimation problem, when an observation Y_k is available the state estimate of $X_{t_{k-1}}$ is updated by the provided information and used for integration of the state to form the prediction of the state X_{t_k}. There is an infinite number of equivalent relations between the input u_t and the output Y_k, the equivalence relation ψ gives a description with the same parameter, but $\tilde{\sigma}$ is independent of z_t.

with the following Bayesian update defining the initial conditions as new observations become available

$$p(\boldsymbol{x}, t_k|\mathcal{Y}_k) = \frac{p(\boldsymbol{y}_k|\boldsymbol{x}_k, \mathcal{Y}_{k-1})p(\boldsymbol{x}, t_k|\mathcal{Y}_{k-1})}{p(\boldsymbol{y}_k|\mathcal{Y}_{k-1})}, \tag{4.9}$$

(4.8) and (4.9) form an iterative procedure, with the probability density function (*pdf*) of the individual observations given by

$$p(\boldsymbol{y}_k|\mathcal{Y}_{k-1}) = \int p(\boldsymbol{y}_k|\boldsymbol{x}_k)p(\boldsymbol{x}, t_k|\mathcal{Y}_{k-1})d\boldsymbol{x}_k, \tag{4.10}$$

and the likelihood given as the product of the individual *pdf*'s, just as in the discrete-time case presented in Chapter 3. The main difference between discrete-time estimation and contonuous-time estimation is (as illustrated in Figure 4.2) that the predictions needs to calculated by integration of the state variables.

4.2 Continuous time Kalman Filtering

The partial differential equation (PDE) (4.8) is quite complex even for seemingly simple SDEs and direct solutions of (4.8) are infeasible. Therefore simulation based methods are relevant in this context. Here we will use an implementation, CTSM[1], (Kristensen & Madsen, 2003; Kristensen et al., 2004b) based on the continuous-time Extended Kalman Filter (EKF). To allow for an efficient implementation of the filter the continuous-discrete time stochastic state-space formulation given in Eqs. (4.3) and (4.4) is restricted to the form

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t; \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{u}_t, t, \boldsymbol{\theta})d\boldsymbol{w}_t \tag{4.11}$$

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{u}_k, t_k; \boldsymbol{\theta}) + \boldsymbol{e}_k \tag{4.12}$$

where $\boldsymbol{\sigma}(\cdot)$ is restricted to the set of quadratic matrix functions ($\boldsymbol{\sigma}(\cdot) \in \mathbb{R}^{n \times n}$), which is independent of the state, and the observation noise ($\boldsymbol{e}_k$) is additive and Gaussian. From (4.8) it is clear that the first restriction ($\boldsymbol{\sigma}(\cdot)$ quadratic) is not a real restriction since the Fokker-Planck equation does not depend on $\boldsymbol{\sigma}(\cdot)$, but on $\boldsymbol{\sigma}(\cdot)\boldsymbol{\sigma}^T(\cdot)$ and the (weak) solution of (4.11) is only unique up to the non-unique "square root" of $\boldsymbol{\sigma}(\cdot)\boldsymbol{\sigma}^T(\cdot)$. This point is illustrated by a small example below.

Example 4.2 *Consider the SDE*

$$dX_t = adt + \sigma_1 dw_{1,t} + \sigma_2 dw_{2,t}, \quad X_0 = 0, \tag{4.13}$$

where a, σ_1 and σ_2 are real constants. The solution to (4.13) is

$$X_t = at + \sigma_1 w_{1,t} + \sigma_2 w_{2,t}, \tag{4.14}$$

which is a Gaussian random variable with mean $E[X_t] = a \cdot t$ and variance $V[X_t] = \sigma_1^2 + \sigma_2^2$ respectively, but this is also the (weak) solution to

$$d\tilde{X}_t = adt + \sqrt{\sigma_1^2 + \sigma_2^2} d\tilde{w}_t, \quad \tilde{X}_0 = 0, \tag{4.15}$$

which illustrates that the weak solution is unique only up to the square root of $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$. The difference between weak and strong solutions is also illustrated here, while X_t and $\tilde{X}_t$ belong to the same distribution these cannot be path-wise equal since they depend on different Brownian motions. □

Actually, for every positive definite symmetric matrix (say $\boldsymbol{\sigma}(\cdot)\boldsymbol{\sigma}^T(\cdot)$) there exists a symmetric positive definite matrix $\tilde{\boldsymbol{\sigma}}(\cdot)$ such that $\tilde{\boldsymbol{\sigma}}^2(\cdot) = \boldsymbol{\sigma}(\cdot)\boldsymbol{\sigma}^T(\cdot)$.

[1]The method is implemented in the open source software CTSM (`www.imm.dtu.dk/~ctsm`)

Even though it might be constructive to think of non-square matrices in the problem formulation, there will always exist a quadratic version which defines the same weak solution.

The second restriction (diffusion independent of the state) is a clear restriction in the multivariate case. To some extent this can be dealt with by adequate transformation as will be discussed in Section 4.3. The last restriction (Gaussian observations) should be dealt with by transformation, if possible. If the sample space is discrete, this can clearly constitute a problem that cannot be dealt with by transformations. Other estimation methods must be considered in such cases.

The continuous-time EKF is very similar to the discrete time EKF stated in Theorem 3.1. The state prediction is, however, replaced by a set of ordinary differential equation (Jazwinski, 1970; Simon, 2006)

$$\frac{d\hat{\boldsymbol{x}}_{t|k}}{dt} = \boldsymbol{f}(\hat{\boldsymbol{x}}_{t|k}, \boldsymbol{u}_t, t, \boldsymbol{\theta}) \quad , \quad t \in [t_k, t_{k+1}) \tag{4.16}$$

$$\frac{d\boldsymbol{P}_{t|k}}{dt} = \boldsymbol{F}\boldsymbol{P}_{t|k} + \boldsymbol{P}_{t|k}\boldsymbol{F}^T + \boldsymbol{\sigma}\boldsymbol{\sigma}^T \quad , \quad t \in [t_k, t_{k+1}), \tag{4.17}$$

where $\boldsymbol{P}_{t|k}$ is the state covariance, and $\boldsymbol{F}$ is the Jacobian of $\boldsymbol{f}$. The state covariance matrix is controlled by a time-homogeneous ODE which enables an explicit solution. However, as both $\boldsymbol{f}(\cdot)$ and $\boldsymbol{\sigma}(\cdot)$ are allowed to depend on time this is a very crude assumption and to increase accuracy the time interval $[t_k, t_{k_1})$ is sub-sampled (Kristensen & Madsen, 2003).

The maximum likelihood estimation is similar to the estimation described in Section 3.1, and the precise formulation of the filter and likelihood estimation are given in Kristensen & Madsen (2003) and Kristensen et al. (2004a). In addition to the predictions and reconstructions given by the filter, CTSM also allows for smoothen state estimates ($\hat{\boldsymbol{x}}_{t|N}$) and statistical test for significance of parameter values. Furthermore, likelihood tests are also available directly since the estimation is based on maximum likelihood estimation.

Figure 4.3 presents a simulation example taken from Paper C and the precise definitions can be found in that reference. The figure illustrates some of the key differences between ODE and SDE estimation. The SDE solution is in this case the smoothened state estimate. It is clear that the SDE solution is able to adapt to the information provided by the observations, while the ODE solution does not adapt to the local information in the state estimate, but only globally when obtaining the best parameter estimate. The ODE residuals are larger and there is a clear autocorrelation seen for the ODE residuals. Furthermore, it turns out that the parameter estimate are also better for the SDE solution than for the ODE-solution.

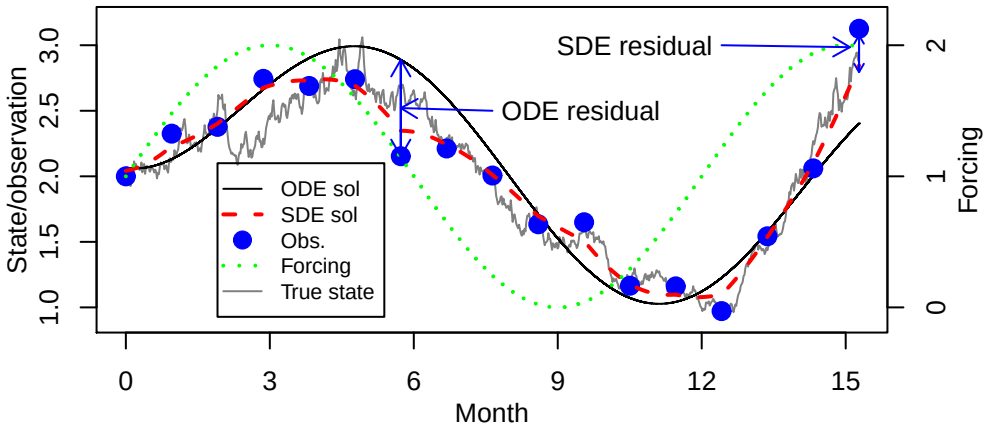

Figure 4.3: Simulation results from the simple simulation example presented in Paper C.

4.3 Transformation of the state-space

As discussed above the filter implemented in CTSM does not allow state dependent diffusion, but if we can find a transformation (see also Figure 4.2)

$$\boldsymbol{z}_t = \boldsymbol{\psi}(\boldsymbol{x}_t, \boldsymbol{u}_t, t; \theta), \tag{4.18}$$

such that $\boldsymbol{z}_t$ is given by the Itô diffusion process

$$d\boldsymbol{z}_t = \tilde{\boldsymbol{f}}(\boldsymbol{x}_t, \boldsymbol{u}_t, t; \boldsymbol{\theta})dt + \tilde{\boldsymbol{\sigma}}(\boldsymbol{u}_t, t; \boldsymbol{\theta})d\boldsymbol{w}_t, \tag{4.19}$$

where the process $\boldsymbol{z}_t$ depends on the same parameters $\boldsymbol{\theta}$ as the process $\boldsymbol{x}_t$, then we can still apply the filter equations defined in CTSM. The results presented in this section is a summary of the results in Papers D and E and details can be found there. Transformations of SDEs is an application of Itô's lemma (the version stated below is credited to Øksendal (2003)).

Theorem 4.1 Itô's lemma: *Let $\boldsymbol{x}_t$ be given by the Itô SDE* (4.3)*. Further, let $\boldsymbol{\psi}(t, \boldsymbol{x})$ be a C^2 map from $\mathbb{R}_0 \times \mathbb{R}^n$ into $\mathbb{R}^q$. Then the process*

$$\boldsymbol{z}_t = \boldsymbol{\psi}(\boldsymbol{x}_t, t), \tag{4.20}$$

is also an Itô process, with the k'th component given by

$$dz_{k,t} = \frac{\partial}{\partial t}\psi_k(\boldsymbol{x}_t, t)dt + \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\psi_k(\boldsymbol{x}_t, t)dx_{i,t} + \frac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j}\psi_k(\boldsymbol{x}, t)dx_{i,t}dx_{j,t}, \tag{4.21}$$

where $dw_{i,t}dw_{j,t} = \delta_{ij}dt$, $dtdt = dw_{i,t}dt = dtdw_{j,t} = 0$.

In the theorem above the dependence of the input $\boldsymbol{u}_t$ is suppressed, but allowed and will act as a time-dependence of $\boldsymbol{\psi}$, i.e. we will need the time derivative of $\boldsymbol{u}_t$ if $\boldsymbol{\psi}$ is a function of the input.

From (4.21) it can be observed that we are looking for a transformation, that solves the set of PDEs

$$\frac{\partial}{\partial x_i}\psi_k(\boldsymbol{x}_t, \boldsymbol{u}_t, t; \boldsymbol{\theta})\sigma_{k,i}(\boldsymbol{x}_t, \boldsymbol{u}_t, t; \boldsymbol{\theta}) = \tilde{\sigma}_{k,i}(\boldsymbol{u}_t, t; \boldsymbol{\theta}). \tag{4.22}$$

Since 4.22 constitues n PDEs in one unknown (ψ_k), it is in general not possible to find a solution to (4.22) (see also Example 7 of Paper D, and the more general treatment in Paper E). Although the pratical aplicability is limited by the existence of an explicit inverse of ψ_k, it is in principle always possible to find the tranformation ψ for univariate diffusion processes.

4.3.1 Univariate diffusion

For univariate diffusion processes it is always possible to find a transformation of the type (4.18). The transformation to unit diffusion is known as the Lamperti transform (Iacus, 2008)

$$\psi(X_t, t) = \int \frac{1}{\sigma(\xi, t)}d\xi\Big|_{\xi=X_t}, \tag{4.23}$$

and the Lamperti transformed process is given by (Luschgy & Pagés, 2006)

$$dZ_t = \left(\psi_t(\psi^{-1}(Z_t, t), t) + \frac{f(\psi^{-1}(Z_t, t), t)}{\sigma(\psi^{-1}(Z_t, t), t)} - \frac{1}{2}\sigma_x(\psi^{-1}(Z_t, t), t)\right)dt + dw_t. \tag{4.24}$$

Even though the Lamperti transform always exist for univariate diffusion processes, the practical application is limited to the existence of an explicit inverse

(Iacus, 2008)

$$x_t = \psi^{-1}(z_t), \tag{4.25}$$

and the implication is most easily illustrated by a small example.

Example 4.3 *Consider the SDE* (4.3) *with the diffusion coefficient* $\boldsymbol{\sigma}$ *given by*

$$\sigma(x_t) = \sigma_0 + \sigma_1\sqrt{x_t}, \tag{4.26}$$

for which the Lamperti transform is

$$\psi(x_t) = \frac{2}{\sigma_1^2}\left(\sigma_1\sqrt{x_t} - \sigma_0 \log\left(\sigma_0 + \sigma_1\sqrt{x_t}\right)\right), \tag{4.27}$$

which does not allow for an explicit inverse function. On the other hand a system with the diffusion coefficient given by

$$\sigma(x_t) = \sqrt{\sigma_0^2 + \sigma_1^2 x_t}, \tag{4.28}$$

has the Lamperti transform

$$\psi(x_t) = \frac{2\sqrt{\sigma_0^2 + \sigma_1^2 x_t}}{\sigma_1^2} \tag{4.29}$$

with the inverse function given by

$$x_t = \psi^{-1}(z_t) = \frac{z_t^2\sigma_1^4 - 4\sigma_0^2}{4\sigma_1^2}. \tag{4.30}$$

In a sense (4.28) *is more natural, because it is equivalent to the diffusion* $\sigma_0 dw_1 + \sigma_1 x_t dw_2$ *(compare with Example 4.2).* □

Paper C presents a simple example of maximum likelihood estimation based on the combination of the EKF and the Lamperti transform. In this example the available water column nitrogen is used as an input for a phytoplankton growth model. In addition to the water-column nitrogen, a time-series of global radiation is available and the resulting simple model is

$$dX_{p,t} = (b_0 U_{w,t} U_{gr,t} - aX_{p,t})dt + \sigma_x X_{p,t} dw_t, \tag{4.31}$$

where $U_{w,t}$ is the available water-column nitrogen and $U_{gr,t}$ is the global radiation at time t. The natural logarithm of the observations are assumed to be

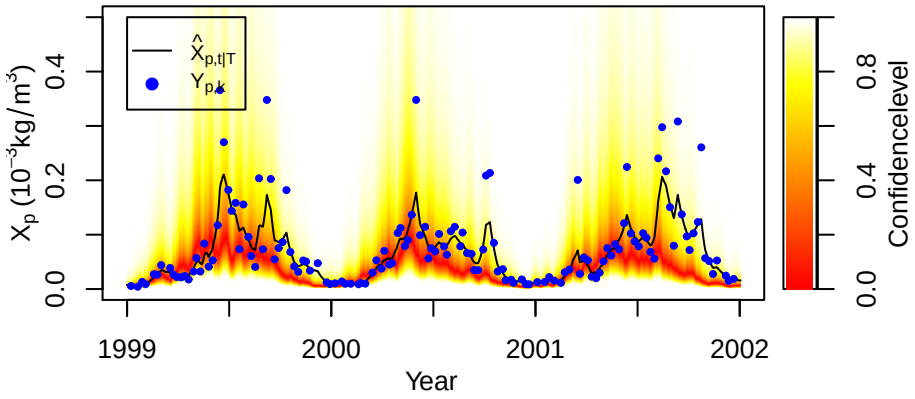

Figure 4.4: Time series of stationary distributions, smoothed state (black line) and observation (blue dots). The colour key refers to confidence intervals around the mode of the stationary distributions.

Gaussian distributed around the natural logarithm of the true state, giving the observation equation

$$\log(Y_{p,k}) = \log(X_{p,t_k}) + e_k, \tag{4.32}$$

where $Y_{p,k}$ is the observed nitrogen content in phytoplankton and $e_k \sim N(0, \sigma_y^2)$. The Lamperti transform is given by $Z_{p,t} = \frac{1}{\sigma_x}\log(X_{p,t})$ and the Lamperti transformed system is given by

$$dZ_{p,t} = \left(\frac{b_0}{\sigma_x}e^{-\sigma_x Z_{p,t}}U_{w,t}U_{gr,t} - \frac{a}{\sigma_x} - \frac{1}{2}\sigma_x\right)dt + dw_t. \tag{4.33}$$

The choice of the diffusion term is simple in the sense that the state-space of $Z_{p,t}$ is $\mathbb{R}$, and the process (4.31) is therefore well suited for a first approach.

The conditional densities are difficult or impossible to find even for simple SDEs (like (4.31)), because it requires the solutions of the PDE given by the Fokker-Planck equation. The stationary distribution is determined by an ODE rather than a PDE ($p_t(\cdot) = 0$) and it might be available, even when the conditional distribution is not. In the case of (4.31), the stationary distribution for fixed $U_{gr,t}$ and $U_{w,t}$ is given by an inverse gamma distribution. It is therefore possible to compare the stationary distribution at each time-point with the observations (Figure 4.4, taken from Paper C). The smoothen state and the mode of the stationary distribution appear to be quite close, even though extreme events are not captured well by the stationary distribution. This is due to the fact that

there are no mechanisms to capture extreme behaviour in the SDE formulation (4.31).

4.3.2 Multivariate diffusion

As has already been illustrated, it is not possible to find Lamperti-type transformations for general multivariate SDEs. Paper E correct an earlier reported result and the resulting diffusion, which allow a Lamperti type transformation, is

$$\boldsymbol{\sigma}(\boldsymbol{x}_t, \boldsymbol{u}_t, t; \boldsymbol{\theta}) = \tilde{\boldsymbol{\sigma}}(\boldsymbol{x}_t, \boldsymbol{u}_t, t; \boldsymbol{\theta}) \boldsymbol{R}(\boldsymbol{u}_t, t; \boldsymbol{\theta}), \tag{4.34}$$

where $\tilde{\boldsymbol{\sigma}}$ is a diagonal matrix with diagonal elements given by $\tilde{\sigma}_{ii}(\boldsymbol{x}_t, \boldsymbol{u}_t, t; \boldsymbol{\theta}) = \tilde{\sigma}_i(x_{i,t}, \boldsymbol{u}_t, t; \boldsymbol{\theta})$, then the SDE admits a Lamperti type transformation (Luschgy & Pagés, 2006). The transformation is given by (Paper D,E)

$$z_t^i = \psi^i(x_t^i, \boldsymbol{u}_t, \boldsymbol{\theta}, t) = \int \frac{d\xi}{\tilde{\sigma}^i(\xi, \boldsymbol{u}_t, \boldsymbol{\theta}, t)} \bigg|_{\xi = x_t^i}, \tag{4.35}$$

and the transformed process is given by

$$dz_t^i = \left(\psi_t(\cdot, t) + \frac{f_i(\cdot)}{\sigma^i(\cdot)} dt - \frac{1}{2} \sigma_x^i(\cdot) \sum_{j=1}^{N} [\boldsymbol{R}(\cdot)]_{i,j}^2 \right) dt + \sum_{j=1}^{N} [\boldsymbol{R}(\cdot)]_{i,j} \, dw_j, \tag{4.36}$$

the proof i provided in Paper D.

The SDE (4.36) does not have unit diffusion and would therefore normally not be referred to as a Lamperti transform. Since time dependence of the diffusion will often be through an observed input, transformation to unit diffusion will involve time-differentiation of observed inputs. The diffusion is therefore intentionally left as a function of the matrix $\boldsymbol{R}(\cdot)$, and a splitting of $\boldsymbol{\sigma}(\cdot)$ such that $\tilde{\boldsymbol{\sigma}}(\cdot)$ is independent of the input is therefore recommended, if possible and reasonable.

While (4.34) is sufficient for the existence of a Lamperti type transformation it is not necessary (Aït-Sahalia, 2008). For instance, some mass balance models, like the one presented in Papers E and G, allow the introduction of Lamperti type transformations by reducing the dimension of the state space.

4.4 Restriction of the state space - Limitations of the Lamperti transform

In general, state dependent diffusion is introduced to give a better description of the transition probabilities compared with the description by models with state-independent diffusion. A central part of this description is the natural restrictions in the state-space of the deterministic skeleton. It is natural to assume that these restrictions are carried over to the SDE description. For population models, a natural restriction of the state-space is the positive real axis and such a restriction implies that the diffusion should tend to zero when the state tend to zero.

From the maximum likelihood estimation point of view presented here, the simplest state-dependent diffusion is state-proportional diffusion, since the state-space of the Lamperti transformed process then becomes the entire real axis (Paper C). As illustrated in Paper F, significant improvements can be achieved by introducing an exponent less than 1 in the state dependence. However, if the exponent is less than (or equal) 0.5, the topology of the state-space changes and estimation by the EKF and Lamperti transformation is no longer possible (or at least likely to fail), because the inverse transformation cannot be evaluated (see also Section 3.1 in Paper E).

The points above will be illustrated by considering the CKLS model (Chan, Karolyi, Longstaff & Sanders, 1992), given by

$$dx_t = (b + ax_t)dt + \sigma x_t^{\gamma} dw_t, \tag{4.37}$$

where $(a, b) \in \mathbb{R}^2$, $\sigma > 0$ and $\gamma > 0$ are constants. In the following we will limit the analysis to situations where $b > 0$ and $a < 0$. The state-space of the deterministic skeleton of the model is a subset of the open interval $(0, \infty)$ depending on the parameters a and b, while the state-space of the stochastic model also depends on the parameters σ and γ. Since $dx_t|_{x_t=0} = 0$ and the Brownian motion is continuous almost surely $(a.s.)$, it is however clear that the state-space most be a subset of the interval $[0, \infty)$, and the critical point is the limit $x_t \to 0$.

For $\gamma = 1$ the Lamperti transform (z_t) of x_t is $z_t = \frac{1}{\sigma}\log(x_t)$ and the state-space of z_t is therefore the real axis, and hence the Lamperti transformed process is given by

$$dz_t = \left(\frac{b}{\sigma x_t} + \frac{a}{\sigma} - \frac{\sigma}{2}\right) dt + dw_t. \tag{4.38}$$

In the limit $x_t \to 0$ we get $dz_t \to \infty$ *a.s*, and z_t can therefore not reach $-\infty$ (corresponding to $x_t = 0$) in finite time. The implication for the filter is that when the transformed state is estimated as being small the predictions will be pushed away from the boundary.

For $\gamma \in (0, 1)$ the Lamperti transform of x_t is given by

$$z_t = \psi(x_t) = \frac{x_t^{1-\gamma}}{\sigma(1-\gamma)} \quad \Rightarrow \quad x_t = \psi^{-1}(z_t) = [\sigma(1-\gamma)z_t]^{\frac{1}{1-\gamma}} \tag{4.39}$$

Since the transformation is monotone, $\psi(0) = 0$, and $\psi(\infty) = \infty$, the state-space of z_t is seen to be a subset of the interval $[0, \infty)$. The Lamperti-transformed process is given by

$$dz_t = \left(\frac{b}{\sigma} x_t^{-\gamma} + \frac{a x_t^{1-\gamma}}{\sigma} - \frac{1}{2}\sigma\gamma x_t^{\gamma-1} \right) dt + dw. \tag{4.40}$$

Restricting the attention to cases where $\gamma > \frac{1}{2}$ and taking the limit $x_t \to 0$, we get $dz_t = \infty$ *a.s.*, and therefore z_t will be pushed away from the boundary independently of the other parameters in the system. On the other hand if $\gamma < \frac{1}{2}$ then $\lim_{x_t \to 0} dz_t = -\infty$ *a.s.*, and z_t will therefore leave the state-space with a positive probability. The problem is that the Lamperti transform is not valid on the boundary of the state-space and if the likelihood of reaching the boundary is too large, the system will reach the boundary in finite time with a positive probability.

The case $\gamma = \frac{1}{2}$ is of special interest and is called the Feller diffusion in biological models and the Cox-Ingersoll-Ross model in finance (Iacus, 2008). The Lamperti transformed process is given by

$$dz_t = \left(\frac{b}{\sigma\sqrt{x_t}} + \frac{a\sqrt{x_t}}{\sigma} - \frac{\sigma}{4\sqrt{x_t}} \right) dt + dw \tag{4.41}$$

$$= \left(\frac{1}{\sigma\sqrt{x_t}} \left(b - \frac{\sigma^2}{4} \right) + \frac{a\sqrt{x_t}}{\sigma} \right) dt + dw, \tag{4.42}$$

where $x_t = \frac{\sigma^2}{4} z_t^2$. The asymptotic behaviour of z_t splits into three cases

$$\lim_{x_t \to 0} dz_t = \begin{cases} +\infty & a.s \quad b > \frac{\sigma^2}{4} \\ dw & \quad\quad b = \frac{\sigma^2}{4} \\ -\infty & a.s \quad b < \frac{\sigma^2}{4}. \end{cases} \tag{4.43}$$

Implying that the transformed process has a positive probability of leaving the state-space when $b < \frac{\sigma^2}{4}$. This is in fact surprising, given that the stationary *pdf* $(p(x))$ of the Feller diffusion (viewed as a function of b) has a discontinuity

in the limit $x_t \to 0$, at $b = \frac{\sigma^2}{2}$ (Feller, 1951). With $p(\lim x \to 0; b > \frac{\sigma^2}{2}) = 0$, $p(\lim x \to 0; b > \frac{\sigma^2}{2}) = \infty$ and $p(\lim x \to 0; b = \frac{\sigma^2}{2}) = c$, where c is a function of a and b. The stationary distribution of the Feller diffusion is the Gamma distribution, with scale and shape parameters depending on a, b, σ and for $b \geq \frac{\sigma^2}{4}$ the Gamma *pdf* is square integrable, while for $b < \frac{\sigma^2}{4}$, it is not. Providing an argument that for $b < \frac{\sigma^2}{4}$, there is simply too much probability mass in the neighbourhood of zero.

Given the reasoning above, the limitations of the Lamperti transform appear at points, where the conditional density is very difficult to determine, because the probability density function assume infinity.

When working with a multivariate nonlinear model (as in Paper F), where the drift b is a function of both time (through a time dependent input) and other states of the system, the easiest way to ensure a consistent formulation of the topology of the assumed state-space (strictly positive states) is by restricting γ to the (open) interval $(\frac{1}{2}, 1)$.

More general structures on the state-space can also be imposed by careful construction of the diffusion term. Paper G analyses a bacteria growth model where the bacteria concentration is restricted by the available substrate and the concentration is therefore restricted to an interval. These kind of restrictions can be imposed by a logistic diffusion term (Schurz, 2007)

$$\sigma(x) = \sigma_0 x^\alpha \left(1 - \frac{x}{K}\right)^\beta, \tag{4.44}$$

where α and β determine the shape of the diffusion function, and K determines the maximum level of x. The Lamperti transform, and an explicit inverse cannot be derived for general α and β, and more importantly for α or β less than or equal $\frac{1}{2}$, we encounter the same problem as illustrated with the CKLS model. In the specific example presented in Paper G, it turns out that the diffusion is much higher when the bacteria concennntration is high than when the bacteria concentration is low (corresponding to $\beta >> \alpha$). To deal with this structure in a framework where the Lamperti transform can be derived, we suggest the following diffusion structure

$$\sigma(x) = \sigma_0 x \left(1 - \left(\frac{x}{K}\right)^\gamma\right), \tag{4.45}$$

where γ is a constant larger than 1. In this case the Lamperti transform and an explicit inverse can be defined and the state-space of the transformed process is $(-\infty, \infty)$. In Paper G, γ is estimated as $>> 1$, implying that the diffusion is close to state-proportional diffusion. However, the formulation (4.45) obeys the natural restrictions of the state-space, while state-proportional diffusion does not (Paper G).

The initial setup of the problem given in Paper G is two-dimensional with the diffusion in each state depending on both states. The formulation is therefore not a subset of the models described in Eq. (4.34). The correlation coefficient of $\boldsymbol{\sigma}(\cdot)\boldsymbol{\sigma}^T(\cdot)$ is however equal to -1 in the presented example, and a reduction of the dimension of the state-space is therefore possible. The resulting diffusion process is therefore essentially one-dimensional, and the analysis illustrates how the Lamperti transform can be applied to a class of models with strict mass balance in the diffusion.

4.5 Maximum likelihood estimation, EKF and the Lamperti-transform

The focus in Section 4.3-4.4 is transformations that allow estimation in the framework presented by Kristensen et al. (2004) (CTSM). The explicit assumption in the reference is that the one step transition probabilities are well approximated by Gaussian distributions. Paper E illustrates that applying the EKF to the Lamperti-transformed process might give much better state estimates than direct application of the EKF (Figure 4.5). The conclussion in Paper E is therefore that even if state dependent diffusion would be allowed in CTSM, it would still be advisable to apply the estimation methodology to the Lamperti transformed process rather than to the original process.

Further Paper E illustrates that reliable parameter estimates can be achieved in multivariate nonlinear SDEs (stochastic Lotka-Volterra systems), also when the diffusion term is dominant. It is, however, also illustrated that the parameter variance provided by CTSM is not reliable when diffusion dominates the system. Still the general conclusion is that maximum likelihood estimation by the EKF and Lamperti transformation is reliable for moderate diffusion and regular sampling. Paper E therefore provide good evidence that the estimation results obtained in Papers F-G (and Section 4.6) can indeed be trusted.

4.6 Model development

The statistical method presented in Section 4.2 provides a framework for estimating the parameters $\boldsymbol{\theta}$, when the structure of $\boldsymbol{f}(\cdot)$, $\boldsymbol{\sigma}(\cdot)$, and $\boldsymbol{h}(\cdot)$ is known. These structures might, however, be unknown and the systematic framework for model improvements presented in this section provides a supplement to traditional hypothesis testing.

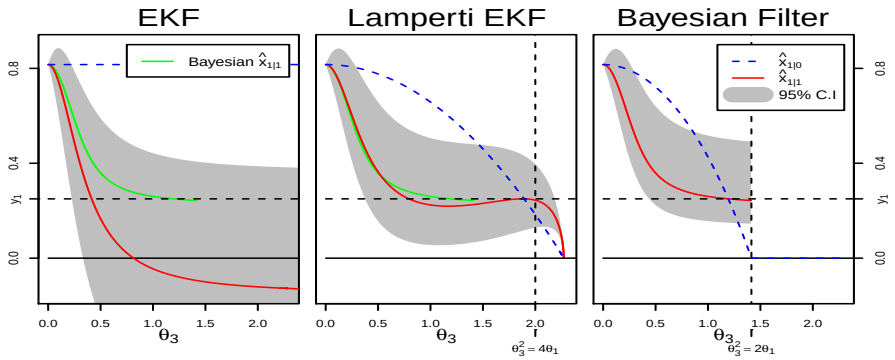

Figure 4.5: State predictions and state reconstructions of the CIR-model with log-normal distributed observations, "EKF" refer to direct application of the EKF, "Lamperti EKF" refer to application of the EKF to the Lamperti-transform process. The estimation details are given in Paper E.

The model development procedure presented in Paper F is a further development of the framework presented in Kristensen et al. (2004b). The statistical framework presented is summarised in Figure 4.6, and this sections will discuss the main steps of the approach.

Step 0:
Step 0 is basically a pre-modelling step and even though a small discussion is provided in Paper F, it is not considered in the presented example where observations are simply assumed to be log-normal distributed (see also the discussion of the special role of the log-normal distribution in Chapter 2). Paper G provides a more thorough discussion of this step. The basic idea is to formulate a minimal model given by

$$dB_t = \mu B_t(T_0 - \eta B_t)dt + \sigma_B B_t(T_0 - \eta B_t)dw, \tag{4.46}$$

where B_t is bacteria concentration, T_0 is the maximum bacteria concentration (measured in substrate units), μ the growth-rate and σ_B the diffusion coefficient. The idea of Paper G is to develop both the drift and the diffusion term to get a more accurate description of data. As a preliminary analysis of data, measured bacteria concentrations were plotted both in the natural and the log domain. These plots indicated that untransformed data would not span the uncertainty of the observation errors when bacteria concentration is low. On the other hand log-transformed data would not span the variation when concentration is high. Based on this reasoning, it was decided to use a Box-Cox transformation (e.g.

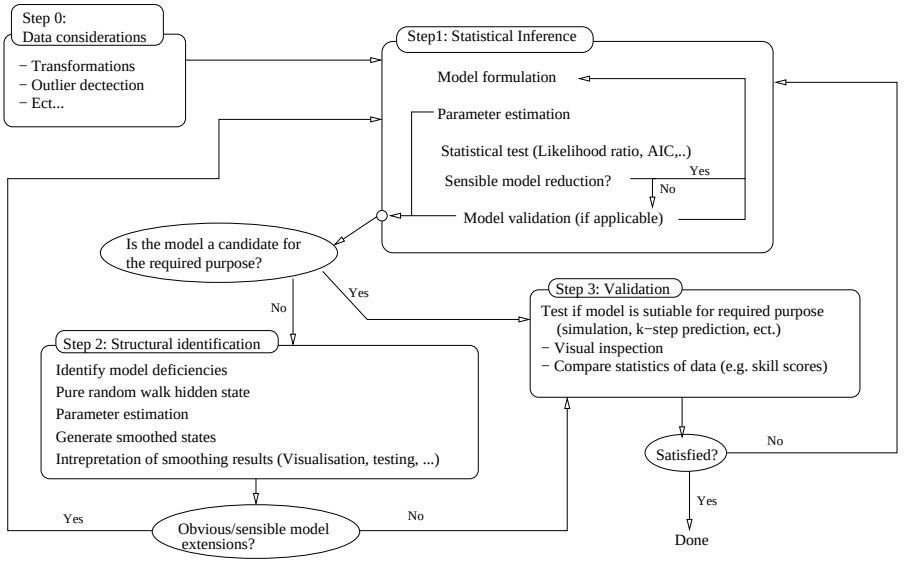

Figure 4.6: Conceptual diagram of suggested model development method.

Madsen, 2008)

$$Y_t^{(p)} = \begin{cases} (Y_t^p - 1)/p & p \neq 0 \\ \log(Y_t) & p = 0. \end{cases} \tag{4.47}$$

Given that the continuous-discrete-time state-space formulation is correct, the standardised one-step prediction errors (the residuals) should be a white noise standard Gaussian random process. The standardisation is with respect to the one step prediction standard deviation, $\sqrt{\Sigma^{yy}_{k|k-1}}$ (since observations are one-dimensional the square root is well defined), see Theorem 3.1 for the discrete time version of $\sqrt{\Sigma^{yy}_{k|k-1}}$. To find the best value of p the model Eqs. (4.46)-(4.47) were fitted for different values of p. Since the system equation (4.46) is assumed to be insufficient, we do not expect uncorrelated residuals. We do however, expect the variance to be approximately constant over time. For low values of p the variance is large in the beginning of the time-series and lower towards the end, while for high values of p the variance is low in the beginning and high towards the end of the series (Figure 4.7). Based on Figure 4.7 and the reasoning about the raw data plots we chose $p = 0.75$.

The analysis above is just one of several data considerations which could be performed. Other relevant issues to adress are outlier detection, data aggregation, etc.

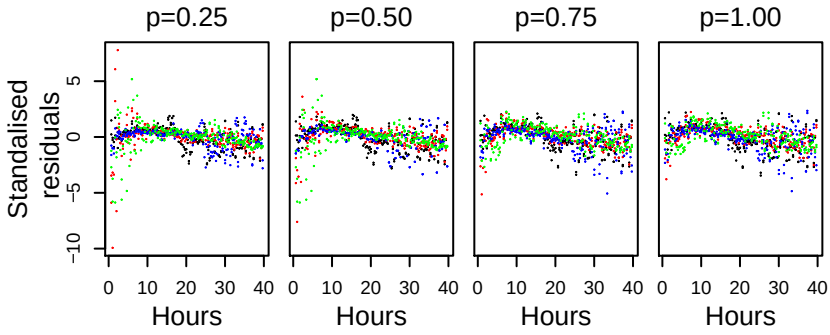

Figure 4.7: Standardised residuals of the model Eqs. (4.46)-(4.47) for different values of p, the colours refer to different experimental setups and bacteria, see Paper G for details.

Step 1 and 2:
Step 1 is the classical statistical inference step and is based on the EKF and maximum likelihood estimation as presented in Section 4.2, with a small discussion provided in Paper F. Step 2 is the first step of the model development procedure, and in step 0 and 1 a simple model is formulated. In this context we think of a simple model as the simplest model that encapsulates the hypothesised linear interactions, while observing the natural restrictions of the state-space (see also Section 4.4). The initial estimation will give an estimate of parameters describing the diffusion ($\boldsymbol{\sigma}(\cdot)$) and the observation variance ($\boldsymbol{S}$), while noting that large values in these matrices pinpoint model deficiencies in the corresponding state. As an example, we consider the initial model setup presented in Paper F (see also Example 4.1)

$$d\begin{bmatrix} X_{w,t} \\ X_{p,t} \end{bmatrix} = \begin{bmatrix} N_{ex,t}Q_t \\ 0 \end{bmatrix} dt + \begin{bmatrix} -Q_t - a_{wp} - a_{wl} & a_{pw} \\ a_{wp} & -a_{pw} - Q_t \end{bmatrix} \begin{bmatrix} X_{w,t} \\ X_{p,t} \end{bmatrix} dt + \begin{bmatrix} \sigma_w X_{w,t} & 0 \\ 0 & \sigma_p X_{p,t} \end{bmatrix} \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} d\boldsymbol{w}_t, \tag{4.48}$$

where $X_{w,t}$ is the water-column nitrogen not contained in phytoplankton and $X_{p,t}$ is the phytoplankton nitrogen. The drift parameters (a_{ij}) are explained in Paper F, $N_{ex,t}$ is the nitrogen input and Q_t is the freshwater input (see also Figure 4.1 and Example 4.1). The available observations are total water-column nitrogen (Y_{TN}), phytoplankton nitrogen (after conversion from chlorophyll, Y_p), and primary production in nitrogen units per day (after conversion from $mgC/m^2/d$, Y_{pp}), under the assumption of log-normal observations the

observation equation is given by

$$\begin{bmatrix} \log(Y_{TN,k}) \\ \log(Y_{p,k}) \\ \log(Y_{pp,k}) \end{bmatrix} = \begin{bmatrix} \log(X_{w,k} + X_{p,t_k}) \\ \log(X_{p,k}) \\ \log(a_{wp} X_{w,k}) \end{bmatrix} + \begin{bmatrix} e_{TN,k} \\ e_{p,k} \\ e_{pp,k} \end{bmatrix}, \tag{4.49}$$

where $e_{i,k}$ are independent white noise sequences with $e_{i,k} \sim N(0, s_i^2)$. From the estimation of this model we obtained a phytoplankton diffusion σ_p, that was about 10 times the water-column nitrogen diffusion σ_w, implying that we should focus on parameters describing phytoplankton dynamics (a_{wp} and a_{pw}). Furthermore the observation noise standard deviation of primary production (s_{pp}) was about 3 times that of phytoplankton (s_p) (which in turn was about 5 times that of total water-column nitrogen (s_{TN})). Combining these results, we focus on the primary production parameter (a_{wp}).

The next step is to formulate an extended state-space, where the parameter (say θ_i) chosen for further inspection is formulated as a random walk hidden state, i.e. we formulate the following system equations for the extended state-space formulation

$$d\begin{bmatrix} \boldsymbol{x}_t \\ \theta_{i,t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}(\boldsymbol{x}_t, \theta_{i,t}, \boldsymbol{u}_t, t; \tilde{\boldsymbol{\theta}}) \\ 0 \end{bmatrix} dt + \begin{bmatrix} \boldsymbol{\sigma}(\boldsymbol{x}_t, \theta_{i,t}, \boldsymbol{u}_t, t; \tilde{\boldsymbol{\theta}}) & \mathbf{0} \\ \mathbf{0} & \tilde{\sigma} \end{bmatrix} d\begin{bmatrix} \boldsymbol{w}_t \\ \tilde{w}_t \end{bmatrix}, \tag{4.50}$$

where $\tilde{\boldsymbol{\theta}} = [(\boldsymbol{\theta} \setminus \theta_i)^T, \theta_{i,0}]^T$, $\tilde{\sigma}$ is the random walk diffusion parameter and $\tilde{w}_t$ is the Brownian motion of the random walk process. The observation equation remains unchanged.

The parameters ($[\tilde{\boldsymbol{\theta}}^T,\ \tilde{\sigma}]^T$) of the extended state-space model is then estimated. The estimated values of the extended state-space is not of great interest, because the model will not be suited for simulation and long term predictions. The smoothen state can, however, suggest possible improvements, if the resulting state estimate of the random walk parameter is plotted as a function of the other state estimates and possible explanatory inputs. The visualisation can be aided by non-parametric methods, like Generalised Additive Models (GAM) (Hastie & Tibshirani, 1990), to get a better visualisation of the covariation of the point cloud. The plot leads to a hypothesis which can be tested in Step 1. The procedure is iterative, and Step 1-2 are repeated until there are no more obvious improvements to the model.

In the example above, the primary production parameter was chosen for further analysis, and the aim is to identify a functional relationship between primary production and the state variables (X_p and X_w), and the input (we only consider global radiation). This is formulated as the function f_i and the primary

production process is given by

$$a^i_{wp,t} = a^0_{wp} X_{w,t} f_i(\boldsymbol{X}_t, \boldsymbol{u}_t), \tag{4.51}$$

where f_i is the functional expression to be identified (with $f_1(\boldsymbol{X}_t, \boldsymbol{u}_t) = 1$ in the linear model). In each step, i, a^0_{wp} is replaced by $a^0_{wp,t}$ to identify a new candidate model, $i+1$ (replacing $f_i(\cdot)$ with $f_{i+1}(\cdot)$). Primary production should be greater than zero and the random walk process is therefore formulated in the log-domain

$$d\log(a^0_{wp,t}) = \sigma_{a_{wp}} dw_{a_{wp},t}. \tag{4.52}$$

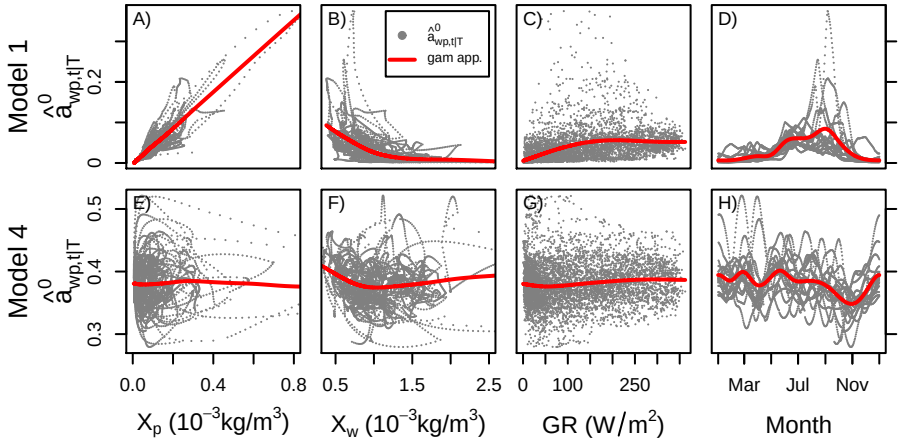

Figure 4.8: Smoothened state of the random walk phytoplankton growth parameter ($\hat{a}^0_{wp,t|T}$) for two different models (Model 1 and 4 of Paper F) as a function of smoothen phytoplankton level (first column), smoothen water column nitrogen (second column), observed global radiation (third column), and time of year (fourth column). Grey dots represent point estimation of $\hat{a}^0_{wp,t|T}$, while red lines represent GAM (smoothing splines in column 1-3 and periodic cubic B-splines in column 4) fits to the points.

The smoothen version of the random walk primary production is plotted as a function of the state estimates and global radiation (Figure 4.8, top row). To improve the visualisation, GAM smoothing splines are fitted to the states and the inputs, and plotted on top of the point cloud, while periodic cubic smoothing splines (see, de Boor (1978) and Møller (2006) for the specific implementation used here), are used to visualise the yearly variation. Clearly the strongest relation is with phytoplankton (Figure 4.8.A) and this was therefore included in the primary production process model.

The procedure is iterative and the steps described above were repeated until there was no obvious model extensions (Figure 4.8, lower column). The individual iterations in the example are described in Paper F. Even though there is some covariation between the primary production parameter and water-column nitrogen and between the primary production parameter and time of year, it is not clear how to incorperate this in the model formulation, and more importantly the correlation is much weaker than for the initial model. Further, phytoplankton diffusion was only about two times that of water column and the observation standard deviation of both primary production and phytoplankton was significantly lower than for the initial model. The parameterisation of the final model identified by the model development steps above is

$$d\begin{bmatrix} X_{w,t} \\ X_{p,t} \end{bmatrix} = \begin{bmatrix} -Q_t - a_{wp}\frac{X_{p,t}Gr_t}{(k_w - X_w)(k_{gr} - Gr_t)} - a_{wl} & a_{pw} \\ a_{wp}\frac{X_{p,t}Gr_t}{(k_w - X_w)(k_{gr} - Gr_t)} & -a_{pw} - Q_t \end{bmatrix} \begin{bmatrix} X_{w,t} \\ X_{p,t} \end{bmatrix} dt + \begin{bmatrix} N_{ex,t}Q_t \\ 0 \end{bmatrix} dt + \begin{bmatrix} \sigma_w X_{w,t} & 0 \\ 0 & \sigma_p X_{p,t} \end{bmatrix} \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} d\boldsymbol{w}_t. \tag{4.53}$$

The model includes nonlinear phytoplankton growth and a nonlinear dependence with global radiation which potentially can explain the seasonality. The model is considered as a candidate for a simulation model and the model is now analysed by the validation in Step 3.

Step 3:
The methodology of the validation step essentially depends on the required purpose of the model. If the required purpose is comparable to maximum likelihood estimation (implying short-term predictions) then the model formulated in Step 1-2 is likely to produce good results. This also implies that reliable validation should be performed on an independent data set in such cases. On the other hand, if the required purpose is different from short-term prediction (e.g. long-term predictions or simulations) then the reuse of data in validation is much less problematic.

Continuing with the example of Paper F the objective was to formulate a model suitable for simulation studies, and as such the model was evaluated with respect to the performance of the simulations. As in Step 2 the initial evaluation was based on visual inspection. The full analysis is given in Paper F, but the most central plots for the argumentation are shown in Figure 4.9, where it is seen that the initial simulation model Eq. (4.53) clearly is not stationary (Figure 4.9.A). An inspection of the phytoplankton drift term led to the introduction of a small constant drift in the phytoplankton level and this turned out to be sufficient to give a stationary simulation where the annual variation from the observations is captured by the simulations (Figure 4.9.B).

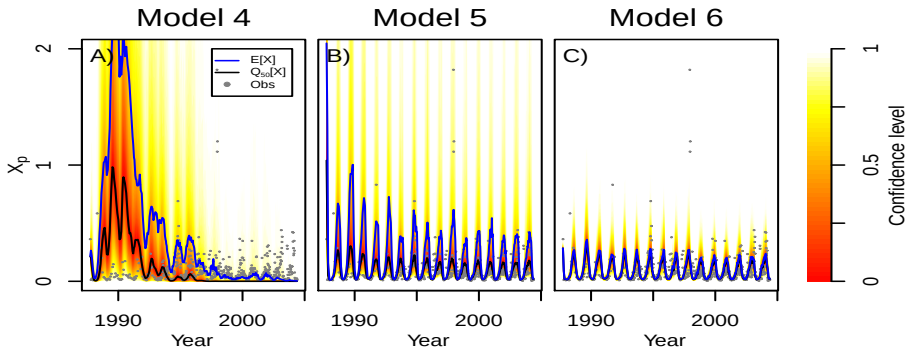

Figure 4.9: Simulation for the time span of observations with Model 4-6, colour code refer to confidence intervals around the median of the distribution, and gray dots are the measurements.

Even though the situation in Figure 4.9.B is much better then the situation in Figure 4.9.A, it is still somewhat unsatisfactory that the distribution contains very heavy tails in the growth season. The large confidence bands imply that the diffusion scales too fast with the state and a weaker dependence can be hypothesised based on Figure 4.9.B, like the hypothesised phytoplankton diffusion term is $\sigma_p X_p^{\gamma_p}$ with $\gamma_p \in (0.5, 1)$. The same hypothesis was applied to the water column diffusion, and the estimation gave good statistics for both γ_p and γ_w which was both significantly different from 1. More importantly the heavy tails in the simulated density disappeared (Figure 4.9.C) and the simulated density of phytoplankton seems more reasonable.

The model development illustrated in Figure 4.9.A-C seems convincing, but evaluation based on visual inspection is clearly subjective by nature and should be confirmed by a more objective skill evaluation. In Paper F objective skill scores were chosen as the total quantile skill score and interval skill score (Gneiting & Raftery, 2007). While the first score evaluates the global (in terms of quantiles) skill, the second evaluates local skill. The conclusion of both scores was that Model 6 (Figure 4.9.C) should be chosen as the final model.

The final step of the modelling exercise was to check for model reductions. The estimation procedure pointed at two reductions ($k_w = 0$ and $r_{12} = 0$). From a mechanistic point of view the first one is unreasonable, since $k_w = 0$ can lead to negative concentrations of water-column nitrogen. This also proved impossible to estimate by CTSM. The second model reduction did not change the likelihood of the model significantly and the final continuous-discrete time

state-space formulation is therefore given by the system equation

$$d\begin{bmatrix} X_{w,t} \\ X_{p,t} \end{bmatrix} = \begin{bmatrix} N_{ex,t}Q_t - (Q_t + a_{wl})X_{w,t} - \frac{a^0_{wp}X_{w,t}X_{p,t}GR_t}{(k_w+X_{w,t})(k_{gr}+GR_t)} + a_{pw}X_{p,t} \\ a_{p0} + \frac{a^0_{wp}X_{w,t}X_{p,t}GR_t}{(k_w+X_{w,t})(k_{gr}+GR_t)} - (a_{pp} + Q_t)X_{p,t} \end{bmatrix} dt + \begin{bmatrix} \sigma_w X_{w,t}^{\gamma_w} & 0 \\ 0 & \sigma_p X_{p,t}^{\gamma_p} \end{bmatrix} d\boldsymbol{w}_t, \tag{4.54}$$

and the observation equation

$$\begin{bmatrix} \log(Y_{TN,k}) \\ \log(Y_{p,k}) \\ \log(Y_{pp,k}) \end{bmatrix} = \begin{bmatrix} \log(X_{w,k} + X_{p,t_k}) \\ \log(X_{p,k}) \\ \log(a_{wp}X_{w,k}) \end{bmatrix} + \begin{bmatrix} e_{TN,k} \\ e_{p,k} \\ e_{pp,k} \end{bmatrix}. \tag{4.55}$$

The final model comprises a more complicated structures than the initial model. The model extensions that were found on the basis of the pure random walk hidden state analysis fit well with established knowledge of primary production, while the extensions based on the evaluation of the simulated densities are less obvious and further γ_i is purely a feature of the stochastic model.

4.7 Discussion

The continuous-discrete time stochastic state-space framework presented in this chapter is based on maximum likelihood estimation, and the likelihood function is evaluated using the Extended Kalman Filter. For nonlinear models Kalman filtering is an approximate method and the idea is to approximate the local solution to the Fokker-Planck equation with Bayesian updating by local Gaussian distributions. The central assumption, in the implementation of the Extended Kalman Filter and the maximum likelihood estimation in CTSM, is therefore that the reconstruction of the conditional probability density (conditioning on information up to present) can be approximated by a Gaussian density, and that the one-step transition density is also Gaussian (possibly after the transformation $\boldsymbol{h}$).

The assumptions above imply that the observation equation should be approximately Gaussian. This in turn implies two assumptions: 1) the observation error is approximately Gaussian, and in CTSM this is ensured by the additive Gaussian observation noise, and 2) the one-step state predictions should be Gaussian. It is relatively straightforward to argue about assumption 1), because it is not influenced by the dynamical behaviour of the system (given that

2 is fulfilled). On the other hand, assumption 2 depends on the sampling intervals: Evidently the transition density of any SDE with continuous $\boldsymbol{f}$ and $\boldsymbol{\sigma}$ will be approximately Gaussian when sampling intervals approach zero ($\boldsymbol{w}_t$ is Gaussian). On the other hand, we have shown that the stationary density of some of the estimated models are very far from Gaussian. To obtain reliable parameter estimates, the sampling intervals must therefore be small enough to allow Gaussian approximations of the one step transition probabilities.

It is shown how a class of state dependent diffusions can be estimated in CTSM by applying the Lamperti transformation. The asymptotic analysis of the Lamperti transformed process shows that we cannot always expect to get stable estimation results from the transformed process in CTSM. It is also shown that the processes excluded by the asymptotic analysis coincide with processes with a very large proportion of the probability density mass cumulated at a single point. These distributions will clearly be poorly approximated by Gaussian distributions, and in this sense the applicability of the Lamperti transform helps us to exclude the most extreme behaviour. As illustrated in Paper E applying the EKF to the Lamperti transformed process might also improve the quality of the Gaussian approximation considerable. Even with these remarks, it still remains to confirm that one-step predictions are Gaussian or to evaluate the quality of the results by other measures. While there exist methods to evaluate the transitions densities (Bak et al., 1999), this approach is neither considered in this summary report nor in any of the Papers C,F-G, where the quality of the results are evaluated by independent assessments.

In Paper C, the estimated path of phytoplankton nitrogen is compared (by visual inspection) with the stationary distribution, which is independent of the local information, clearly it depends globally on the observations through the maximum likelihood estimation of parameters. In Paper F, observations are compared with simulation confidence intervals by visual inspection as well as quantitative methods, while model selection is based on both maximum likelihood estimation and quantile skill scores. In Paper G, both standardised residuals and simulation plots are used to evaluate the quality of the presented models.

The significant role of the particular parametric formulation of the diffusion term is illustrated in Papers F and G. In both cases large improvements in the likelihood were obtained by developing the diffusion term. While the improvements were mostly on the bandwidth of the confidence bands of the simulated distribution in Paper F, the dynamics of the model changed completely by introducing a modified logistic diffusion term in Paper G.

The model development framework presented in Section 4.6 and Paper F, worked well in the presented example, and the developed model is consistent with con-

ventional NP-models like the one presented in Fasham et al. (1990). The model setup in Paper F is quite simple and contains only two states. This choice makes the model appropriate for demonstrating the methodology, because we only need to focus on the parameterisation of the primary production process. The methodology is, however, not limited to simple models, but for more complex models the model analysis would be more complicated, in particular the first selection might be difficult because there might be several parameter which are obvious choices for further investigation.

Investigation of regime shift models like the ones presented in Chapter 2 is further complicated by the facts, that measurements of eelgrass are sparse and strictly non Gaussian (many zero observations), furether regime shifts do not occur often and there might therefore be very few point available to infer the parameters controlling the shifts.

Chapter 5

Conclusion

The central theme in this thesis is the formulation of stochastic dynamical systems. The first part (Chapter 2) present simple dynamical systems for ecological regime shifts in discrete time. The presented simulation studies emphasise the key importance for the stochastic perturbations and how these will affect the system dynamics. System noise in dynamical systems is often thought of as additive and Gaussian with no impact on the mean dynamics. This assumption is however not true for nonlinear systems and in particular for regime shift model, because the random dynamics can push the system into an alternative stable state.

The introduction of noise in an additive Gaussian framework might also lead to unacceptable state estimate or simulation results (like negative concentrations). In order to deal with such phenomena, the random perturbations need to be state dependent. A simple approach is to include the multiplicative equivalent to the additive Gaussian noise. This leads to multiplicative log-normal noise on the state, and it is shown how this deals effectively with the problem of negative states. The strong impact of noise formulation is illustrated in a very simple setup and emphasise the importance of a proper stochastic formulation.

State and parameter estimation is a central part of stochastic modelling. In Chapter 3, a method is suggested, for parameter and state estimation in models with multiplicative log-normal noise. The main contribution in this chapter is

the direct formulation of the filter equations for multiplicative noise and the likelihood formulation under the hypothesis of log-normal distributed multiplicative noise. Direct state estimates cannot be used in maximum likelihood estimation, because the resulting covariance matrices might not be admissible for the multivariate log-normal distribution. Therefore a covariance matrix transformation which ensure admissibility is presented.

Continuous time stochastic modelling (here SDEs) offers a more intuitive model formulation of both the deterministic and the stochastic parts of the model. State estimation offers the opportunity of formulating data driven model improvements. The main contribution in Chapter 4 is the model development framework, where a combination of pure random walk hidden states and investigation of simulated distributions are used for model development in a simple stochastic differential equation NP-model. In addition, the importance of a proper description of the diffusion term is stressed.

The general applicability of the model development framework is illustrated in Paper G, where a restricted bacterial growth model is analysed. In the restricted SDE model the importance of a proper description of the diffusion term becomes very clear, because the natural restriction is not just positive states but an interval. Further, the importance of a proper diffusion term is illustrated by the change of the key stoichiometric constant, when a more adequate description of the diffusion is implemented.

Bibliography

Aït-Sahalia Y. (2008) Closed-form likelihood expansions for multivariate diffusions. *The Annals of Statistics*, **32**, (2), 906-937

Bak, J., Nielsen, H.Aa., and Madsen, H. (1999) Goodness of fit od stochastic differential equations. *Symposium i Anvendt Statistik, Copenhagen*, 341-346

Carpenter S.R., Ludwig D. and Brock W.A. (1999) Management of Eutrophication for Lakes Subject to Potentially Irreversible Changes. *Ecological Applications*, **9**(3), 751-771

Carpenter S.R. (2005) Eutrophication of Aquatic Ecosystems: Bistability an Soil Phosphorus. *PNAS*, **102**(29), 10002-10005

Chan K.C., Karolyi G.A., Longstaff F.A. and Sanders A.B. (1992) An empirical investigation of alternative models of the short term interest rate. *The Journal of Finance*, **47**(3), 1209-1227

de Boor C. (1978) A practical Guide to Splines. *Applied mathematical sciences*; 27. Springer-Verlag.

Fasham, M.J.R, Ducklow, H.W., and McKelvie S.M. (1990) A nitrogen-based model of plankton dynamics in the oceanic mixed layer. *Journal of Marine Research*, **58**, 591-639

Feller W. (1951) Two singular diffusion problems. *The Annals of Mathematics, Second Series*, **54**(1), 173-182

Fenton L.F. (1960) The Sum of Log-Normal Probability Distributions in Scatter Transmission Systems. *IEEE Transaction and Communication*, **8**(1), 57-67

Gard T.C. (1988) Introduction to stochastic differential equations. *Marcel Dekker, Inc.*, New York.

Givon D., Stinis P. and Weare J. (2009) Variance reduction for Particle filters of Systems With Time Scale Separation. *IEEE Transactions on Signal Processing*, **57**, (2), 424-435

Gneiting T., and Raftery A.E. (2007) Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, **102**(477), 359-378

Hastie T.J., and Tibshirani R.J. (1990) Generalized additive models. *Monographs on Statistics and Applied Probability*, **43**, Chapman & Hall/CRC

Holling C.S. (1973) Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, **4**, 1-23

Iacus S.M. (2008) Simulation and Inference for Stochastic Differential Equations - With R Examples. *Springer Series in Statistics*

Jazwinski A.H. (1970) Stochastic Processes and Filtering Theory. *Academic Press*, New York.

Jimenez J.C., and Ozaki T. (2002) Linear Estimation of Continuous-Discrete Linear State Space Models with Multiplicative Noise. *System & Control Letters*, **47**, 91-101

Kalman R.E. (1960) A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME-Journal of Basic Engineering, Series D*, **82**, 35-45

Karatzas I. and Shreve S.E. (1991) Brownian Motion and Stochastic Calculus - Second edition. *Springer-Verlag, New York.*

Klebaner F.C. (2005) Introduction to Stochastic Calculus with Applications. *Imperial College Press*

Kotz S., Balakrishnan N., and Johnson L. N. (2000) Continuous Multivariate Distributions, Volume 1: Models and Applications. *Wiley Series in Probability and Statistics*

Kristensen N.R., and Madsen H. (2003) Continuous time stochastic modeling - CTSM 2.3 - Mathematics Guide. *Technical University of Denmark*

Kristensen N.R., Madsen H., and Jørgensen S.B. (2004) Parameter estimation in stochastic grey-box models. *Automatica*, **40**, 225-237

Kristensen N.R., Madsen H., and Jørgensen S.B. (2004b) A method for systematic improvement of stochastic grey-box models. *Computers & Chemical Engineering*, **28**, 1431-1449

Limpert E., Stahel W.A., and Abbt M. (2001) Log-normal Distributions across the Sciences: Keys and Clues. *BioScience*, **51**(5), 341-352

Luschgy H., and Pagés G. (2006) Functional quantization of a class of Brownian diffusions: A constructive approach. *Stochastic Processes and their Applications* **116**, *310-336*

Madsen, H. (2008) Time Series Analysis. *Chapman & Hall*

May R.M. (1977) Thresholds and breakpoints in ecosystems with multiplicity of stable states. *Nature*, **269**, 471-477

Møller J.K. (2006) Modeling of Uncertainty in Wind Energy Forecast. *Master Thesis Informatics and Mathematical Modelling, Technical University of Denmark.* URL:http://www.imm.dtu.dk/pubdb/p.php?4428.

Møller J.K., Carstensen J., and Madsen H. (2006) Regime Shift Models for Simulation of the Interaction between Benthic and Pelagic Production. *IMM-Technical Report-2006-23*, http://orbit.dtu.dk/Publications,resultSetTable.recordLink.sdirect?sp=200972

Møller J.K., Carstensen J., and Madsen H. (2009) Parameter Estimation in State Space Models with Multiplicative Noise - Examples. *IMM-Technical Report-2009-09* Informatics and Mathematical Modelling, Technical University of Denmark URL: http://orbit.dtu.dk/All.external?recid=240477

Scheffer M., Szabó S.,van Nes E.H., Rinaldi S., Kautsky N., Norberg J., Roijackers R.M.M., and Franken R.J.M. (2003) Floating Plant Dominance as a Stable State. *PNAS*, **100**(7), 4040-4045

Schindler D.W. (2006) Recent advances in the understanding and management of eutrophication. *Limnology and Oceanography*, **51**(1), 356-363

Schurz H. (2007) Modelling, analysis and discretization of stochastic logistic equations. *International journal of numerical analysis and modelling*, **4**(2), 178-197

Simon D. (2006) Optimal stats esitmation- Kalman, H_∞, and nonlinear approaches. *John Wiley & Sons, Inc.*

Tier C., and Hanson F.B. (1981) Persistence in density denpendent stochastic popurlations *Mathematical Biosciences*, **53**, 89-117

Tong H. (1990) Non-linear Time Series - A Dynamical System Approach. *Oxford Science University Press*

Zhang L., and Zhang X.D. (2007) An Optimal Filtering Algorithm for Systems With Multiplicative/Additive Noises. *IEEE Signal Processing Letters*, **14**(7),469-472

Øksendal B. (2003) Stochastic Differential Equations - An Introduction with Applications, Sixth edition. *Springer, Berlin.*

Part II

Papers

Paper A

Dynamic two state stochastic models for ecological regime shifts

Authors:
J.K. Møller, J. Carstensen, H. Madsen, and T. Andersen.

Published in:
Environmetrics, **20**, 912-927, (2009).

Dynamic Two State Stochastic Models for Ecological Regime Shifts

Jan Kloppenborg Møller[1,2], Jacob Carstensen[2],
Henrik Madsen[1] and Tom Andersen[3]

Abstact

A simple non-linear stochastic two state, discrete-time model is presented. The interaction between benthic and pelagic vegetation in aquatic ecosystems subject to changing external nutrient loading is described by the nonlinear functions. The dynamical behavior of the deterministic part of the model illustrates that hysteresis effect and regime shifts can be obtained for a limited range of parameter values only. The effect of multiplicative noise components entering at different levels of the model is presented and discussed. Including noise leads to very different results on the stability of regimes, depending on how the noise propagates through the system. The dynamical properties of a system should therefore be described through propagation of the state distributions rather than the state means and consequently, stochastic models should be compared in a probabilistic framework.

Key Words: Aquatic ecosystems, bistability, hysteresis, stochastic state space models, multiplicative noise.

1 Introduction

Eutrophication is a result of nutrient enrichment of ecosystems. For aquatic ecosystems the link between algae blooms and eutrophication was recognized in the 1960s (Schindler (2006)). The main nutrient sources causing eutrophication are generally nitrogen (N) in marine systems and phosphorus (P) in freshwater systems. The human contribution to nutrient inputs derives from industry and

[1]DTU Informatics, Richard Pedersens Plads, Technical University of Denmark Building 321, DK-2800 Lyngby, Denmark.
[2]National Environmental Research Institute, Fredriksborgvej 399, DK-4000 Roskilde, Denmark.
[3]University of Oslo, Biologisk Institut, Kristine Bonevies hus, Blindernveien 31, 0371 Oslo, Norway.

households (dominant sources of P), as well as agriculture (dominant source of N).

Holling (1973) and May (1977) are early examples of analyses of systems with multiple stable states in biology. In more recent years this hypothesis has also been proposed for aquatic systems (see Schindler (2006) for a historic presentation). The presence of alternative stable states is potentially reflected in hysteresis effects or even irreversible changes of a system caused by excessive nutrient loading (eutrophication) on the system. There is a bulk literature on this subject in the context of lakes, e.g. Scheffer et al. (2003), Carpenter (2005), and others. These kinds of shifts, where the dynamics of the system change suddenly, will be referred to as regime shifts.

An example of a simple system with regime shifts is given in Carpenter et al. (1999), where P in the water column is modeled as a function of the loadings. The dynamic part includes a non-linear recycling term. This construction can give rise to alternative stable states, hysteresis effects and irreversible changes can be observed.

Others (e.g. Scheffer et al. (2003)) give examples of two dimensional (e.g. floating and submerged plants) systems with bistable dynamics. Scheffer et al. (2003) offer a conceptual empirical verification of the bistable hypothesis in controlled experiments. Such verification is, however, difficult in real ecosystems. The presence of different time constants in the system may (as pointed out by Carpenter (2005)) appear to be hysteresis effects or even irreversible changes of the system. Systems may take hundreds or even thousands of years to recover after heavy loading has caused a break down of the system over a matter just of years or decades. Therefore,it can also be argued that, from a practical point of view, the system has experienced an irreversible change.

An important issue for systems with alternative stable states is the stability of the states. Such analysis can be addressed with bifurcation analysis (see e.g. Scheffer et al. (2003)), and this will show how far the system is from an abrupt change. Since ecosystems are very complex systems that cannot be encapsulated in a simple mathematical formulation, there is bound to be noise in the system ("system noise"). In addition to this, measurements of variables related to these systems are uncertain. An adequate description of these systems is therefore a stochastic one. The stochastic nature of the system implies that we can only assign a probability of a regime shift appearing in the next time interval, and the best strategy might therefore be to use a precautionary policy.

Ludwig et al. (2003) and Carpenter et al. (1999) support this approach and they explore management strategies for lakes using an economic utility model added to the dynamical system description. Both papers advocate a precaution-

ary policy due to the uncertainty in the system. However, studies examining the effect of stochastic perturbations in aquatic models are still few and far between, and mostly concerned with stochastic variations in model forcing. For example, Ludwig et al. (2003) analyzed the sensitivity of a simple lake model having two states, phosphorus in the water column and in the sediments respectively, with respect to a log-normal distributed phosphorus loading. Carpenter & Brock (2006) analyze a full stochastic model in the SDE (stochastic differential equations) framework, this analysis led to a proposed management strategy based on variance monitoring.

The objective of this study is to illustrate hysteresis and regime shifts in a simple model consisting of two states for pelagic and benthic vegetation respectively. Moreover, we will analyze the model with stochastic perturbations added to the forcing as well as in the system equations. Guttal & Jayaprakash (2007) analyze the effect of stochastic forcing in two one-dimensional systems with hysteresis effects and deduce some results on the assymptotic behavior of these systems. We will analyze a two-dimensional system and further we will focus more on the transients. The difference between observation noise and system noise will also be emphasized. The adequate description of such systems is a state space formulation.

We will present some characteristics of this model and emphasize where caution is needed to ensure stability. The model is formulated in a discrete time setting and not, as is often the case, by means of differential equations. The state space model and its rationale is introduced in Section 2. The behavior of the deterministic part of the model is analyzed in Section 3. In Section 4 the effect of stochastic perturbations added to model input and system equations is investigated.

2 The State Space Model

Aquatic systems undergoing the adverse effects of eutrophication are characterized by changes from benthic to pelagic production (Scheffer et al. (2001), Carpenter (2005)). Here, we formulate a general model for aquatic ecosystems with two states for a nutrient (typically nitrogen for marine ecosystems and phosphorus for lakes), and analyze it within a state space framework. This implies that we will separate the random variations into "system noise" and "observation noise", in addition noise in the input is allowed. The aim is to analyze how the different noise components propagate through the system of

equations. The stochastic state space model is given by the system equation

$$\begin{bmatrix} X_{p,t} \\ X_{b,t} \end{bmatrix} = N_{ex,t} \begin{bmatrix} a_p \\ 0 \end{bmatrix} + \begin{bmatrix} \xi_{p,t} & 0 \\ 0 & \xi_{b,t} \end{bmatrix} \times \left(\begin{bmatrix} b_{p,t-1} & 0 \\ b_{p,b} f_{t-1} & b_b \end{bmatrix} \begin{bmatrix} X_{p,t-1} \\ X_{b,t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ K f_{t-1} \end{bmatrix} \right), \tag{1}$$

and the observation equation

$$\begin{bmatrix} Y_{p,t} \\ Y_{b,t} \end{bmatrix} = \begin{bmatrix} \lambda_{p,t} & 0 \\ 0 & \lambda_{b,t} \end{bmatrix} \begin{bmatrix} X_{p,t} \\ X_{b,t} \end{bmatrix}, \tag{2}$$

where $X_{p,t}$ and $X_{b,t}$ are the state variables describing the nutrient content in the water column ($X_{p,t}$) and in the benthic vegetation ($X_{b,t}$), at time t. The model is forced by the external nutrient loading ($N_{ex,t}$), which can be a measured or estimated time series, this can be a random or a deterministic function. $\boldsymbol{Y}_t = [Y_{p,t} \quad Y_{b,t}]^T$ is the observed state of the system. a_p is a conversion factor from nutrient input to concentrations in the water column, and $b_{p,t}$ and b_b are the proportions of pelagic and benthic biomass, respectively, carried over from one year to the next. For the water column this retention factor $b_{p,t}$ is a time-varying parameter that depends on the benthic vegetation (see below), whereas the parameter b_b is time-invariant. $b_{p,b} f_t$ is the proportion of nutrients from the water column which is used for growth in the benthic region. The nutrient content in the sediments is not modeled explicitly, but is implicitly contained in $b_{p,b} f_t$ and $b_{p,t-1}$. $K f_t$ describes a background level of benthic vegetation, e.g. immigration of seeds from outside the system.

$\{\xi_{j,t}, \lambda_{j,t}\}$ (with $j \in \{p, b\}$) are mutually independent noise terms. For simplicity additive normal distributed noise is commonly considered in the literature, but as we require the state variables to be positive we have chosen multiplicative noise with $P(noise < 0) = 0$. This ensures that the observations, and the states remain positive at all times. We further require that all the random variables have expectation equal to one and are mutually independent white noise sequences, i.e. sequences of iid. random variables.

The nonlinear effects in the model are the functions $b_{p,t} = b_p(X_{b,t})$ and $f_t = f(X_{p,t})$. $b_{p,t}$ is the recycling of nutrient from the sediment to the pelagic region, assuming that it is low when the benthic vegetation is high and high when the benthic vegetation is low. This functional relationship is represented by a sigmoid function:

$$b_{p,t} = (1 - \Phi(X_{b,t}))d + t_p = (1 - \Phi(X_{b,t}; r_p, \tau_p^2))d + t_p \tag{3}$$

where $\Phi(x; r_p, \tau_p^2)$ is the CDF for a Gaussian random variable with mean r_p and variance τ_p^2. This function generates a regime shift and potentially a hysteresis

Table 1: Parameters of the model

Parameter	a_p	$b_{p,b}$	b_b	K	d	t_p	r_p	τ_p	r_b	τ_b
Value	0.8	0.1	0.8	0.2	0.4	0.2	0.4	0.1	0.4	0.1

effect. If $\tau_p^2 = 0$ we will refer to this as a hard threshold, otherwise we will refer to this as a smooth threshold. For the limit $\tau_p^2 \to \infty$, $b_{p,t}$ is a constant. The function is constrained between t_p and $t_p + d$.

The function f_t describes the shadowing effect of nutrients in the water column, mainly through particulate organic forms such as phytoplankton and detritus, and is modeled as

$$f_t = 1 - \Phi(X_{p,t}; r_b, \tau_b^2) \tag{4}$$

resulting in values ranging from no shading at all ($f_t = 1$) to complete shading of the benthic vegetation caused by nutrients in the water column ($f_t = 0$). These are asymptotic values ($X_{p,t} \to \pm\infty$), but if τ_b^2 is small then this will be effective values for $X_{p,t}$ small or large.

A short-hand notation of the state space model is

$$\boldsymbol{X}_t = N_{ex,t}\boldsymbol{a} + \boldsymbol{\Xi}_t(\boldsymbol{B}_{t-1}\boldsymbol{X}_{t-1} + \boldsymbol{k}_{t-1}) \tag{5}$$

$$= N_{ex,t}\boldsymbol{a} + \boldsymbol{\Xi}_t g(\boldsymbol{X}_{t-1}) \tag{6}$$

$$\boldsymbol{Y}_t = \boldsymbol{\Lambda}_t \boldsymbol{X}_t, \tag{7}$$

where the matrices are defined by comparison with Eqs. (1)- (2)

If the noise terms are held constant at one then dynamics of the system are contained entirely in the system equation ($\boldsymbol{X}_t$). Even without stochastic perturbations, model simulations may result in complicated structures.

3 The Deterministic Model

The purpose of this section is to examine the system in Eqs. (5), with the noise term held constant (i.e. $V[\Xi_t] = 0$), and the external loading deterministic ($V[N_{ex,t}] = 0$). One important feature of the model is the hysteresis and associated threshold effect. Thus, the model should be able to describe such features with an appropriate parameterization.

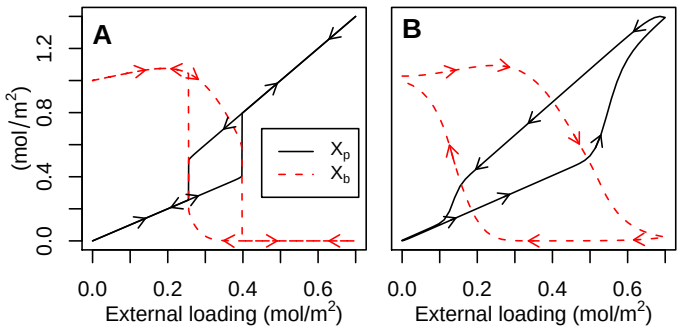

Figure 1: A) Shows stationary points for the system in Eq. (5)-(7) with parameters as in Table 1. B) Shows the trajectory with a loading increased from 0 to 0.7 over a period of 50 years and then decreased from 0.7 to 0 over a period of 50 years.

The stationary solutions to the model (i.e. points where $\boldsymbol{X}_t = \boldsymbol{X}_{t-1}$) were found for a selected set of parameter values (Table 1) over a wide range of nutrient inputs as forcing (left panel of Figure 1). The stationary solutions clearly illustrate the hysteresis effect with a sudden change in system variables (from a system dominated by benthic vegetation to a system with high pelagic primary production) occurring when the nutrient loading exceeds 0.4. Following this regime shift the nutrient input to the system must be decreased to below 0.25 for the system to return to the original state. The right-hand side of Figure 1 shows a simulated response of the system with a linearly increasing nutrient load followed by a linearly decreasing nutrient load. The system response is more gradual and the hysteresis appears to occur over a wider range of nutrient input, this happens because the system is not allowed to reach the stationary point before the loading is changed. Since constant nutrient loading in ecosystems is a hypothetical case seldom observed in practice, this illustrates some of the difficulties with identifying such systems.

For a nutrient input in the midrange between the two threshold values (i.e. between 0.25 and 0.4) the system will converge towards one of the stability points representing the two different regimes (Figure 2). The border line between the two basins of attractions (the seperatrix) is important for studying an intervention of the system. Figure 2 shows two plots linked to the stability of the processes described above. The left panel of the figure shows the norm of $\boldsymbol{X}_t - \boldsymbol{X}_{t-1}$ for $N_{ex,t} = 0.35$ as a function of $\boldsymbol{X}_{t-1}$. The plot shows that there are three equilibria (this is expected from Figure 1). The figure further shows the

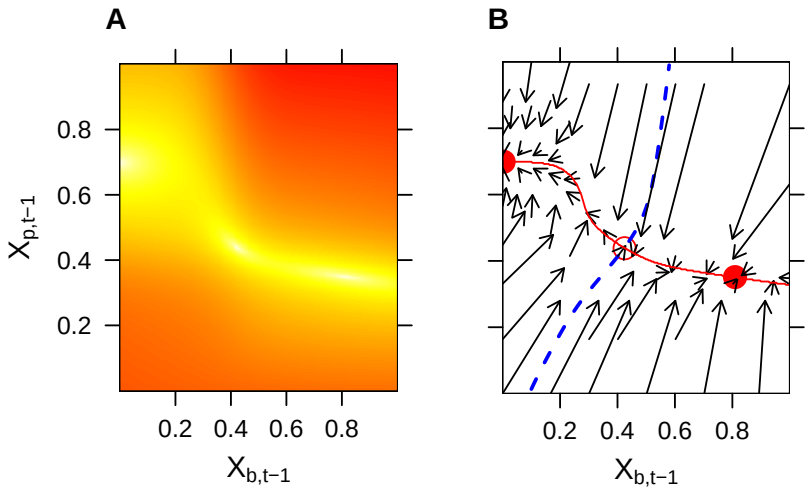

Figure 2: A) Contour plot of $||\boldsymbol{X}_t - \boldsymbol{X}_{t-1}||^2$, with a constant nutrient input, $N_{ex} = 0.35$, and parameters as in Table 1. B) The corresponding three equilibria; two of these are stable (filled circles) and one is unstable (open circle). The system trajectories towards the equilibrium are indicated by length and direction of the vectors $\boldsymbol{X}_t - \boldsymbol{X}_{t-1}$ (arrows). The solid line indicates the minimum of $||\boldsymbol{X}_t - \boldsymbol{X}_{t-1}||^2$ for each X_b, and the dashed line indicates the border of the two basins of attractions (the seperatrix), i.e. sequences starting to the left of this line will converge to the stable equilibrium point indicated with the filled circle to the left and starting points to the right of the dashed line will converge to the filled circle (stable equilibrium point) to the right.

line of minimum difference connecting these three points. The left panel of the figure shows the direction and length of the same vector, in addition the three equilibrium points are indicated and the blue line indicates the seperatrix. van Nes & Scheffer (2007) propose using the recovery rate from small perturbation as an indicator of how close the system is to a break down, this indicator would be the length of the arrows in Figure 2.

Plots like Figure 2 can help us answer questions about the system. E.g., if we are in the left equilibrium point of Figure 2, will it be possible to shift to the other state by removing pelagic biomass from the system? And how much biomass should be removed? A shift from the pelagic dominated state to the benthic dominated state is only feasible provided that $X_b > 0.1$. This may help explain why some bio-manipulation experiments are successful whereas others are not! Another important question that can be addressed with the model is: How much should the loading to the system be changed to cause a regime shift, and how will the transition occur, i.e. will the transition be slow and gradual

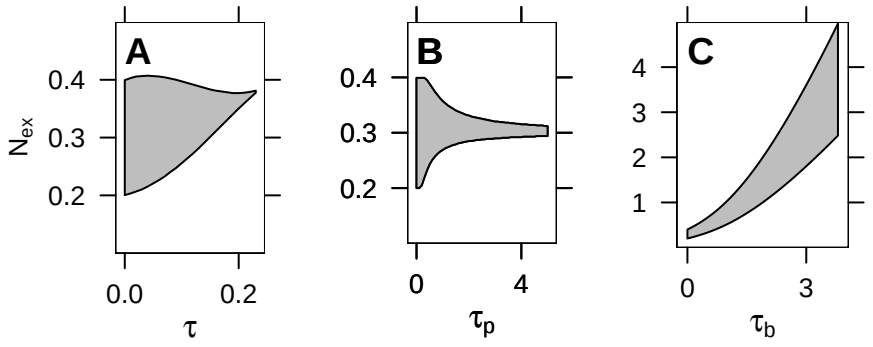

Figure 3: Occurrence of bistability for different nutrient inputs (N_{ex}) and values of the "threshold" parameters: A) $\tau = \tau_p = \tau_b$, B) $\tau_b = 0$ and varying τ_p, and C) $\tau_p = 0$ and varying τ_b. All other parameters are as in Table 1. Note the different scales of the axes.

or more abrupt? Such plots can also illustrate the risk of changing regimes in the more realistic situation when stochastic variation is added to the model.

3.1 Regions of Bistability

The region of bistability varies with the parameter values. As stated above, the model will be linear if we set $\tau_b^2 = \tau_p^2 = \infty$ and bistability will not be present. In a regime shift context, this asymptotic behavior is not very informative, since it only explains that there can be no hysteresis effect. It is more important to investigate when bistability actually occurs. Figure 3 shows some combinations of τ's and nutrient load that lead to bistability when all other parameters are held constant. If $\tau_p = \tau_b = \tau$, it is only possible to obtain a bistable system for τ less than about 0.25 (3A). If either τ_p or τ_b is set equal to zero (hard threshold) then a bistable region exists for a larger (possible infinite) range of values for the other threshold parameter (3B,C). It should also be stressed that there are large differences in the bistability regions, depending on which of the τ^2-parameters is set to zero.

4 The Stochastic Model

Models are abstractions of reality simplifying the complexity of ecosystems into a mathematical tool suitable for addressing specific questions, and as such models are bound to be uncertain. To address the inherent stochastic nature of the system, stochastic models are required. Stochastic perturbations can have large, and sometimes counter-intuitive, impacts on the model behavior. The different uncertainty components described in the introduction propagate differently through the models, and consequently the impacts on model results are different.

The most simple noise component is observation noise, the measurement errors do not propagate through the system equations and therefore the result is random fluctuations added to the deterministic model. System noise on the other hand is mediated through the equations that may lead to complicated correlation structures. Input noise is also filtered through the system equations, but in a more simple way. The difference between input and system noise is discussed in more detail in Appendix A.

4.1 Noisy Input

In the following we will let the loading be a random variable, as was the case for the states and the observation, we will also require this to be positive. This is also ensured by using positive multiplicative noise on the input. I.e. we set

$$N_{ex,t} = \hat{N}_{ex,t} \cdot u_t \tag{8}$$

where u_t is a white noise process with expectation equal to one and $P(u_t < 0) = 0$, and $\hat{N}_{ex,t}$ is a deterministic function, e.g. an estimated sequence of loadings.

Ludwig et al. (2003) used multiplicative $\log t$ distributed noise as input. The $\log t$ distribution does, however, not have a finite expectation, e.g. if $u \sim \log t$ and we would like to compute the expectation of $\boldsymbol{X}_{t|t-1}$ ($\boldsymbol{X}_t$ given $\boldsymbol{X}_{t-1}$), then the expected pelagic nutrient content would be

$$\begin{aligned} E[X_{p,t|t-1}] =& E[N_{ex,t} a_p u_t] + E[\xi_{p,t} b_{p,t-1} X_{p,t-1}] \\ =& N_{ex,t} a_p E[u_t] + b_{p,t-1} X_{p,t-1} E[\xi_{p,t}] \\ =& \infty + b_{p,t-1} X_{p,t-1} = \infty. \end{aligned} \tag{9}$$

The full derivation of the non-existence of the expectation of the $\log t$ distribution is given in Appendix B. It could be argued that for all practical purposes

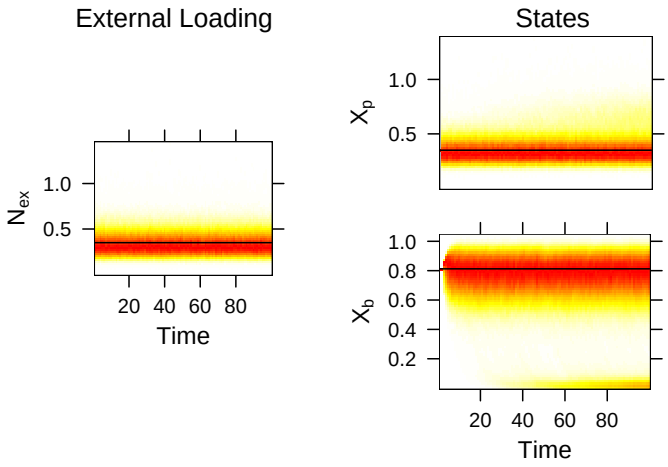

Figure 4: Responses in state variables in a model with input noisy ($V[u_t] = 0.1$) and a constant mean $\hat{N}_{ex,t} = 0.35$, no system noise is added. Solid lines indicate the deterministic response. The noise was simulated by means of a log-normal distribution. The density plots are based on 5000 realizations of the system and are estimated with a Gaussian kernel (the default in "R").

(e.g. simulations) this would not be relevant since the probability of extreme events is still small. It is however unfortunate that we are not able to make the most simple inferences about future states. The log t distribution would further force us to look at quantiles only. For our modeling purposes we want the expectation of the noise terms to equal 1. Consequently, the $\log t$ distribution, as employed by Ludwig et al. (2003), cannot be used and the log-normal distribution is considered a better alternative.

Guttal & Jayaprakash (2007) analyze the asymptotic behavior of two one-dimensional bistable ecological models under stochastic forcing. Their analysis shows that the bistable regime is narrowed by the introduction of noise and that for sufficiently large noise there is not two distinct regimes, but a probability distribution for being in one or the other regime. Forcing in biological systems is bound to vary in time and we will focus more on the transient behavior of the process than on the asymptotic behavior. The analysis carried out here will also point to the inclusion of noise in the system equation, this is important when forcing only has direct effect on one level of the system equations.

Simulations with noise on the input show that there is an increasing probability of shifting between regimes as time passes (see Figure 4). Although the simulations start in the "healthy" state, there is no guarantee that the system will

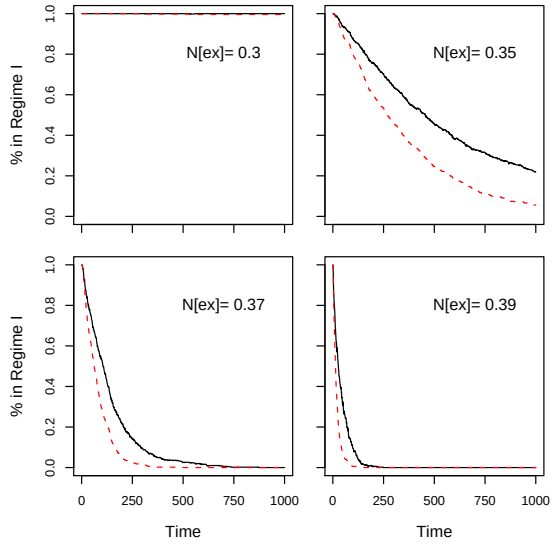

Figure 5: The proportion of sample paths which are in the "healthy" regime (R_1) at time t (solid line), and the proportion of sample paths which have not visited the "unhealthy" regime (R_2) at a time prior to t (dashed line). The plots are based on 500 simulations, of trajectories for $t \in [0, 1000]$.

remain in this state. In fact, the "unhealthy" state is actually absorbing or at least what could be called "effectively absorbing". This means that when the system is in the "unhealthy" state, the probability of returning to the "healthy" state is zero, or at least very close to zero. The onset of this feature is to some extent indicated in Figure 4, but the true nature of the phenomenon reveals itself only after longer simulations (not shown).

We now formally define the two distinct regimes ("healthy" and "unhealthy") for the stochastic model. To do this let $\boldsymbol{X}_{s,I}$ and $\boldsymbol{X}_{s,II}$ be the two stable equilibrium points and let $s(X_b)$ be the separator function:

$$s(x_b) = \{x_p \in \mathbb{R}_0 : \{\boldsymbol{X}_t^{(x_p-\epsilon,x_b)}\} \to X_{s,I}, \{\boldsymbol{X}_t^{(x_p+\epsilon,x_b)}\} \to X_{s,II}, \\ t \to \infty, \forall \epsilon > 0\} \tag{10}$$

where $\{\boldsymbol{X}_t^{\boldsymbol{y}}\}$ is the deterministic process starting at the point $\boldsymbol{y}$, and $s(x_b)$ is the dashed line in Figure 2. The two regimes are now defined as

$$R_1 = \{(x_b, x_p) \in \mathbb{R}_0^2 : s(x_b) > x_p\} \tag{11}$$

$$R_2 = \{(x_b, x_p) \in \mathbb{R}_0^2 : s(x_b) \leq x_p\}. \tag{12}$$

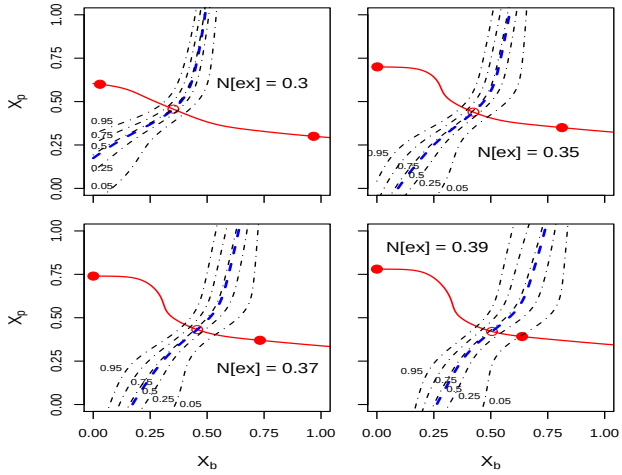

Figure 6: Probability of being in regime II in the next time step as a function of the state variables with four different loadings subject to input noise ($V[u_t] = 0.1$ and $u_t \sim LN$), no system noise.

Thus, R_1 is the "healthy" regime with low phytoplankton levels and high benthic vegetation, and R_2 is the "unhealthy" regime with high phytoplankton levels and low benthic vegetation.

Figure 4 indicates that regime II could be completely absorbing for $\hat{N}_{ex,t} = 0.35$, i.e. corresponding to $P(X_{b,t} \in R_1, X_{b,t-1} \in R_2) = 0$ and $P(X_{b,t} \in R_2, X_{b,t-1} \in R_2) = 1$. This implies that $P(X_{b,t} \in R_2) = P(\exists s \in \{1, 2, ..., t\} : X_s \in R_2)$, and estimates of these two probabilities are plotted in Figure 5. There are sample paths that visit regime "II" and then return to regime "I". However, it is also clear that regime II is absorbing in the long run for the loading $N_{ex,t} = 0.35$, for higher loadings the first exit time is accelerated and all trajectories end up in regime "II", while for the loading $N_{ex,t} = 0.3$ all trajectories stay in regime "I" for the timespan considered (1000 years). For $N_{ex,t} = 0.35$ the system has not reached the stationary distribution and it is possible that a proportion of the trajectories will remain in regime I in the long run.

In stochastic analysis, focus is primarily on the stationary distributions, but the time before the system settles in the steady state can be very long (Figure 5). Loadings are not constant in time since these are subject to natural variations, changes in land use, and political decisions to mitigate eutrophication. It is therefore relevant to look at the transient behavior of the system, since environmental policies should have an effect within a reasonable time horizon. It is

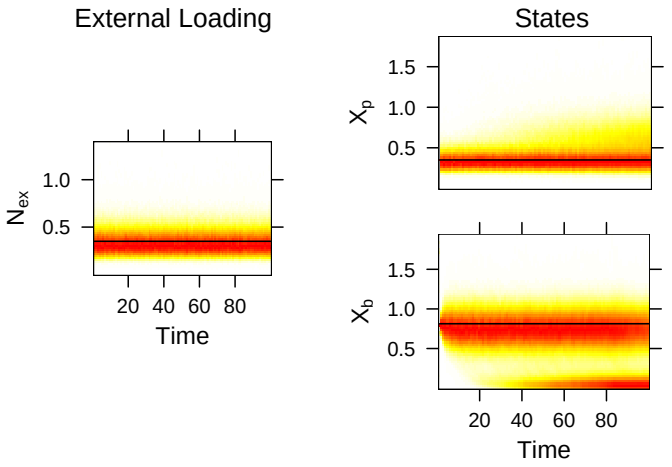

Figure 7: System with input noise and system noise. The solid lines are the deterministic response. Noise was simulated by a log-normal distribution with $V[u_t] = 0.1$, $V[\xi_p] = 0.05$, $V[\xi_b] = 0.01$ and $\hat{N}_{ex,t} = 0.35$ for all t. The density plots are based on 5000 realizations of the system and estimated with a Gaussian kernel (the default in "R").

e.g. not relevant for the policy making that a change in loading will lead to a healthy system in 1000 years.

4.1.1 A One Step Distribution

When the only noise component is on the input we can actually calculate the probability of moving to (or staying in) regime I during the next time step, given the state. Let F_u be the distribution function of u_t then the probability is

$$
\begin{aligned}
P(\boldsymbol{X}_t \in R_2 | \boldsymbol{X}_{t-1}) &= P(X_{p,t} > s(X_{b,t|t-1}) | \boldsymbol{X}_{t-1}) && (13)\\
&= P(u_t \hat{N}_{ex,t} a_p > s(X_{b,t|t-1}) - b_p(X_{b,t-1}) X_{p,t-1}) && (14)\\
&= P\left(u_t > \frac{s(X_{b,t|t-1}) - b_p(X_{b,t}) X_{p,t-1}}{\hat{N}_{ex,t} a_p} \right) && (15)\\
&= 1 - F_u \left(\frac{s(X_{b,t|t-1}) - b_p(X_{b,t-1}) X_{p,t-1}}{\hat{N}_{ex,t} a_p} \right) && (16)\\
&= 1 - F_u(g(\boldsymbol{X}_{t-1})). && (17)
\end{aligned}
$$

$X_{b,t|t-1}$ is a deterministic function of $\boldsymbol{X}_{t-1}$ and the distribution of u is given. $s(X_{b,t|t-1})$ is determined by numerical methods. By solving the equation $g(\boldsymbol{X}_{t-1}) = F_u^{-1}(1-p)$ we can find the functions that describe the probability of moving to regime II during the next time step (this was also done by numerical methods). Solutions of this are shown for a number of different p values and for different loadings in Figure 6. This figure should be compared with Figure 5, and it is seen that even though the probability of moving from $\mathbf{X}_{S,I}$ to regime II in one step is very small ($<< 5\%$), it can still behave like an absorbing state, as is the case for e.g. $N_{ex,t} = 0.37$. The probability of leaving regime I given that the process starts in $\boldsymbol{X}_{S,I}$ and the loading is 0.3 is very close to zero. For a loading of 0.35 we see that the trajectories are slowly leaving regime I, and further we see that most of the trajectories that have left regime I stay in regime II.

From the plots in Figures 5 and 6 we can conclude that regime II is what we can call effectively absorbing. It is however important to stress that the analysis so far only included noise in the loading, this noise has a direct effect on the pelagic level and only an indirect effect on the benthic level.

4.2 System Noise

As discussed earlier, system noise will propagate through the system in a much more complicated way than input noise. The multiplicative noise at time t will be multiplied with system noise at all prior times (see Appendix A). As seen in the previous section, input noise is able to drive the system, in the sense that the system with input noise is very different from a system that fluctuates around the deterministic system (corresponding to observational noise only). As we will see adding noise in different levels in the system will also have very different implications. Noise in the pelagic level propagates through the system in a similar way as input noise, while benthic noise alters the system in a quite different way.

Figure 7 shows a simulation of the model with system and input noise, it is seen that the decay to regime II is much faster than for a system with only input noise (Figure 4). Further it is seen that the spread in the state variables is larger in this system. Quantifying the deviation between two stochastic systems is much more involved than quantifying the deviation between two deterministic systems. In deterministic systems it is enough to follow the trajectory of the state variables and measure how fast these diverge. For stochastic systems we are actually interested in the divergence distribution, i.e. how the distribution evolves.

Chan & Tong (2001) propose an invariant measure for the divergence of distri-

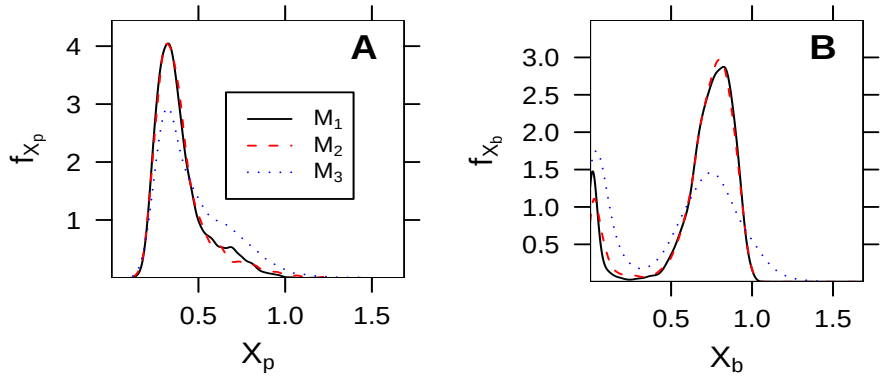

Figure 8: Estimated conditional densities after 100 years, all processes start at the point $\boldsymbol{X}_{S,I}$ with a loading of $N_{ex,t} = 0.35$. M_1 is a model with only noise in the input ($V[u_t] = 0.1$), M_2 is the density with noise in the input and in the pelagic level ($V[\xi_p] = 0.05$). Finally M_3 is the density for a model with noise in the input, the pelagic level and in the benthic level ($V[\xi_b] = 0.01$).

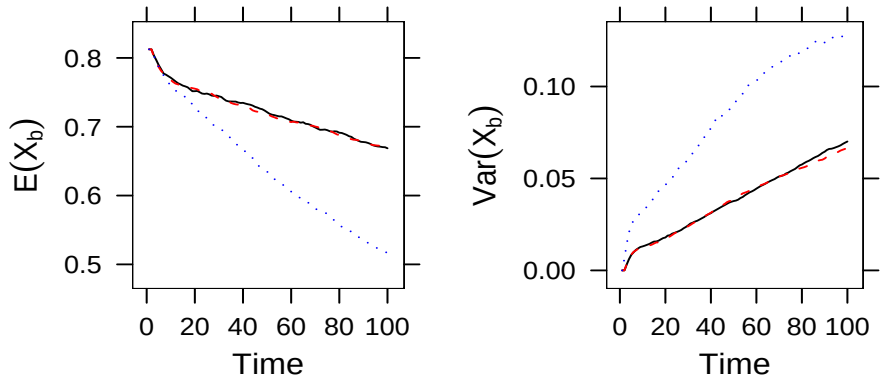

Figure 9: Estimated conditional (on $\boldsymbol{X}_0$) mean values and variances as a function of time, all processes start i $\boldsymbol{X}_{S,I}$ with a loading of $N_{ex,t} = 0.35$. The solid line is a model with only noise in the input ($V[u] = 0.1$) the dashed line is the mean/variance with noise in the input and in the pelagic level ($V[\xi_p] = 0.05$), and the dotted line is the mean/variance for a model with noise in the input, the pelagic level and in the benthic level ($V[\xi_b] = 0.01$).

butions. However, it seems that this requires a very large number of simulations. We will therefore focus on differences in the first exit-times and the mean value and variances of trajectories and note that these can actually be equal for different distributions. It is clear, however, that two distributions with different moments or exit-times are indeed different.

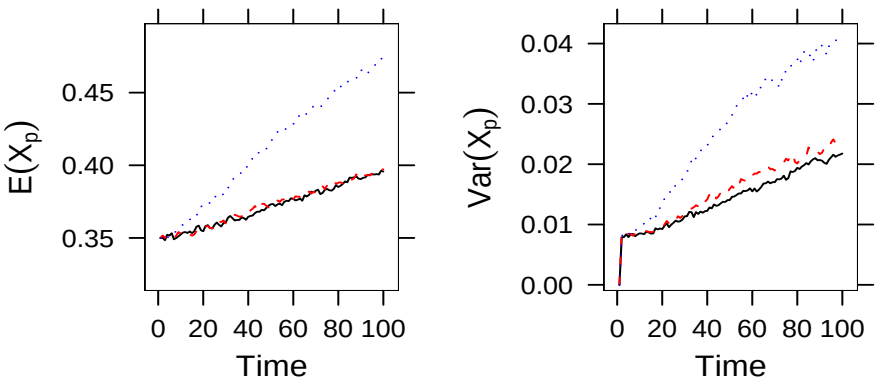

Figure 10: Estimated conditional (on $\boldsymbol{X}_0$) mean values and variances as a function of time, all processes start at $\boldsymbol{X}_{S,I}$ with a loading of $N_{ex,t} = 0.35$ for the deterministic model. The solid line is a model with only noise in the input ($V[u] = 0.1$), the dashed line is the mean/variance with noise in the input and in the pelagic level ($V[\xi_p] = 0.05$) and the dotted line is the mean/variance for a model with noise in the input, the pelagic level and in the benthic level ($V[\xi_b] = 0.01$).

Figure 8 shows the estimated densities for models with noise in different levels of the model, it is seen that adding noise in the pelagic level does not change the conditional distribution much compared to the model with noise only in the input. In contrast adding noise in the benthic level changes the model significantly, with the probability mass being much more evenly distributed over the range of possible values. It is worth noting that this happens even though the variance of the benthic noise is small compared to the noise in the other levels.

Figures 9 respectively 10 show the mean value and variance for X_b respectively X_p, as a function of time. It is seen that both the mean value and the variances diverge for the model with noise in all levels, while the model with only noise in the input and the model with noise in the input and in the pelagic level are close at all considered timepoints. The conclusion of this plot is therefore the same as for Figure 8.

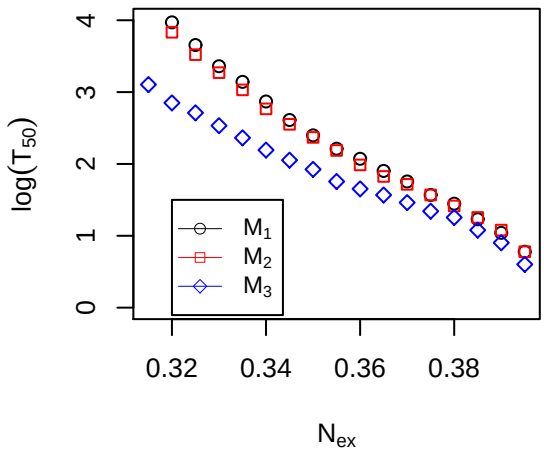

Figure 11: Estimated first exit-time as a function of loading, all processes start at $\boldsymbol{X_0} = \boldsymbol{X}_{S,I}$ with loading of $N_{ex,t} = 0.35$ for the deterministic model. M_1 is the model with only noise in the input ($V[u] = 0.1$), M_2 is the exit time with noise in the input and in the pelagic level ($V[\xi_p] = 0.05$), and M_3 is the exit time for a model with noise in the input, the pelagic level and in the benthic level ($V[\xi_b] = 0.01$).

Figure 11 shows the time before half of the trajectories have left the healthy region given that $\boldsymbol{X}_0 = \boldsymbol{X}_{S,I}$. The conclusion from this plot is essentially the same as for the mean and variance plots, namely that input noise and pelagic noise propagate through the system in very similar ways, while noise in the benthic level changes the system in a quite different way.

Even though moments and exit-times do not give the full information on the system, it seems that we will be able to discriminate the distributions on the basis of these statistics.

5 Conclusion and Discussion

We have presented a nonlinear two state stochastic model for studying the interaction between pelagic and benthic production as a function of nutrient loading.

It has been shown how this model can give rise to bistability and hysteresis effects, the regions of bistability are shown to comprise a small subset of the parameter space. Since there will be uncertainties in the parameter estimates, a consequence of this observation is that it is hard to prove the existence of hysteresis or bistability, as it has to be documented that the parameters belong to the region of bistability for the model.

A main topic in the analysis is the propagation of noise. It is illustrated that noise can be an important driver for the dynamics in the system. Guttal & Jayaprakash (2007) analyze the asymptotic behavior of systems under stochastic forcing, their findings are similar to ours, namely that break down of systems can be accelerated by noise. It is therefore important to consider noise within the system rather than only analyzing the deterministic skeleton of the system and perceive noise as a random fluctuation around the deterministic trajectories. We analyze a two-dimensional system rather than a one-dimensional system and distinction between system noise and input noise is then more important. The focus in this analysis is more on densities in the transient than the asymptotic behavior of the systems.

In order to analyze a stochastic model we need to look at the probabilistic properties of the model and the propagation of densities rather than the trajectories of the deterministic model. Comparing densities is, however, not a simple task, and this will typically require a very large number of simulations, hence we suggest comparing simpler statistics such as the mean and the variance of the (time-varying) densities.

In the setting of aquatic system, these points imply that small fluctuations in the pelagic production may increase the probability of a system collapse towards complete dominance of pelagic production. Once the system has collapsed, it may be very difficult to return the system to the original state. It is therefore important to include stochastic components in analysis of such systems and manage aquatic ecosystems on the basis of a precautionary principle.

van Nes & Scheffer (2007) and Carpenter & Brock (2006) advocate two different approaches to identifying regime-shifts before they actually occur, another approach is to estimate parameters in models like the one presented in this article and then base predictions of densities on simulations studies.

6 Acknowledgments

This study is a contribution to the EU-funded project Thresholds (GOCE-003933). The study was supported by the Danish Graduate School in Biostatistics.

References

Carpenter S. R., Ludwig D. and Brock W. A. (1999) Management of Eutrophication for Lakes Subject to Potentially Irreversible Changes. *Ecological Applications*, **9**(3), 751-771

Carpenter S. R. (2005) Eutrophication of Aquatic Ecosystems: Bistability an Soil Phosphorus. *PNAS*, **102**(29), 10002-10005

Carpenter S. R., Brock W. A. (2005) Rising variance: a leading indicator of ecological transition. *Ecology Letters*, **9**, 311-318

Chan, K.-S., Tong H. (2001) Chaos: A Statistical Perspective. *Springer-Verlag, New York, Inc.*

Guttal V., Jayaprakash C. (2007) Impact of noise on bistable ecological systems. *Ecological Moddeling*, **201**(3), 420-428

Holling, C. S. (1973) Resilience and stability of ecological systems *Annual Review of Ecology and Systematics*, **4**, 1-23

Ludwig D., Carpenter S. R. and Brock W. A. (2003) Optimal Phosphorus Loading for a Potentially Eutrophic Lake. *Ecological Applications*, **13**(4), 1135-1152

May, R. M. (1977) Thresholds and breakpoints in ecosystems with multiplicity of stable states *Nature*, **269**, 471-477

van Nes E. H. and Scheffer M. (2007) Slow Recovery from Perturbations as a Generic Indicator of a Nearby Catastrophic Shift *The American Naturalist*, **169**(6), 738-747

Scheffer M., Carpenter S., Foley J.A., Folke C. and Walker B. (2001) Catastrophic shifts in ecosystems. *Nature*, **413**, 591-596

Scheffer M., Szabó S.,van Nes E. H., Rinaldi S., Kautsky N., Norberg J., Roijackers R. M. M. and Franken R. J. M. (2003) Floating Plant Dominance as a Stable State. *PNAS*, **100**(7), 4040-4045

Schindler D. W. (2006) Recent advances in the understanding and management of eutrophication. *Limnology and Oceanography*, **51**(1), 356-363

A Derivation of Noise Filtration

The purpose of this section is to derive theoretical expressions of how noise at different levels of the state space model is filtered through the system. The state space model is given (as in Section 2) by

$$\begin{aligned} N_{ex,t} =& \hat{N}_{ex,t} \cdot u_t && \text{(A.1)} \\ \boldsymbol{X}_t =& N_{ex,t}\boldsymbol{a} + \boldsymbol{\Xi}_t(\boldsymbol{B}_{t-1}\boldsymbol{X}_{t-1} + \boldsymbol{k}_{t-1}) && \text{(A.2)} \\ =& N_{ex,t}\boldsymbol{a} + \boldsymbol{\Xi}_t g(\boldsymbol{X}_{t-1}) && \text{(A.3)} \\ \boldsymbol{Y}_t =& \boldsymbol{\Lambda}_t \boldsymbol{X}_t && \text{(A.4)} \end{aligned}$$

the aim here is to give an expression of $\boldsymbol{X}_t$ given $\boldsymbol{X}_0$ ($\boldsymbol{X}_{t|0}$) and a sequence of predicted loadings $\{\hat{N}_{ex,i}\}_{i=1}^t$. To do this we use induction of $\hat{\boldsymbol{X}}_t$. The result is

$$\begin{aligned} \boldsymbol{X}_{t|0} =& \boldsymbol{a}\hat{N}_{ex,t}u_t + \boldsymbol{\Xi}_t g(\boldsymbol{X}_{t-1|0}) && \text{(A.5)} \\ =& \boldsymbol{a}\hat{N}_{ex,t}u_t + \boldsymbol{\Xi}_t(B_{t-1}\boldsymbol{X}_{t-1|0} + \boldsymbol{k}_{t-1}) && \text{(A.6)} \\ =& \boldsymbol{a}\hat{N}_{t,ex}u_t + \boldsymbol{\Xi}_t\boldsymbol{k}_{t-1} + \\ & \boldsymbol{\Xi}_t\boldsymbol{B}_{t-1}(\boldsymbol{a}N_{ex,t-1}u_{t-1} + \boldsymbol{\Xi}_{t-1}(\boldsymbol{B}_{t-2}\boldsymbol{X}_{t-2|0} + \boldsymbol{k}_{t-2})) && \text{(A.7)} \\ =& \boldsymbol{a}\hat{N}_{ex,t}u_t + \boldsymbol{\Xi}_t\boldsymbol{B}_{t-1}\boldsymbol{a}\hat{N}_{ex,t-1}u_{t-1} + \boldsymbol{\Xi}_t\boldsymbol{k}_{t-1} + \\ & \boldsymbol{\Xi}_t\boldsymbol{B}_{t-1}\boldsymbol{\Xi}_{t-1}\boldsymbol{k}_{t-2} + \boldsymbol{\Xi}_t\boldsymbol{B}_{t-1}\boldsymbol{\Xi}_{t-1}\boldsymbol{B}_{t-2}\boldsymbol{X}_{t-2|0} && \text{(A.8)} \\ & \vdots \\ =& \sum_{i=0}^{t}(\prod_{j=i}^{t-1}\boldsymbol{\Xi}_{j+1}\boldsymbol{B}_j)\hat{N}_{ex,i}u_i\boldsymbol{a} + \boldsymbol{\Xi}_t\sum_{i=1}^{t}(\prod_{j=i}^{t-1}\boldsymbol{\Xi}_j\boldsymbol{B}_j)\boldsymbol{k}_{i-1} + \\ & (\prod_{i=1}^{t}\boldsymbol{\Xi}_i\boldsymbol{B}_{i-1})\boldsymbol{x}_0, && \text{(A.9)} \end{aligned}$$

where $\prod_{i=k}^{l} A_i = I$ for $k > l$. With this derivation it is seen that noise at different levels will filter differently through the system. The most simple noise i.e. the observation noise, does not influence the dynamics of the system. On the contrary noise in the input is filtered by products of $\boldsymbol{B}_s$'s which are themselves nonlinear functions of the state variables. Finally system noise will filter with products of the $\boldsymbol{B}_s$'s and itself. Even if the system is linear ($\boldsymbol{B}_s = \boldsymbol{B}$ and $\boldsymbol{k}_s = \boldsymbol{k}$ for all s), then the complexity introduced by the multiplicative noise would make direct derivation of expectations and variances very complicated. We will therefore rely on the simulation studies presented in this article.

B The Log t-distribution

In this section we will present the derivations needed to prove that the $\log t$-distribution does not have an expectation. The pdf of a t-distributed random variable with $\nu > 0$ degrees of freedom is

$$f(x) = \frac{\Gamma[(\nu+1)/2]}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2} \tag{B.10}$$

Now if X follows a t-distribution then $Y = e^X$ follows a $\log t$-distributed and we can calculate the expectation of Y as

$$E[Y] = \int_{-\infty}^{\infty} y(x) f(x) dx \tag{B.11}$$

$$= \int_{-\infty}^{\infty} \frac{e^x \Gamma[(\nu+1)/2]}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2} dx \tag{B.12}$$

$$= \frac{\Gamma[(\nu+1)/2]}{\sqrt{\nu\pi}\Gamma(\nu/2)} \int_{-\infty}^{\infty} \frac{e^x}{(1 + x^2/\nu)^{(\nu+1)/2}} dx. \tag{B.13}$$

The denominator under the integral is a polynomial of degree $\nu + 1$ while the enumerator is a polynomial of degree ∞ and we can conclude that $E[Y] = \infty$.

Paper B

Kalman Filtering and Parameter Estimation in State Space Models with Multiplicative Noise

---

Authors:
J.K. Møller, H. Madsen, and J. Carstensen.

Submitted to:
IEEE Transactions on Signal Processing, (2010).

Kalman Filtering and Parameter Estimation in State Space Models with Multiplicative Noise

Jan Kloppenborg Møller[1,2], Henrik Madsen[1] and Jacob Carstensen[2]

Abstact

State space models generally describe the error terms as either additive or multiplicative noise which is then transformed into additive noise by means of the log-transformation, but approaches for state space models are not readily available when the system equations contain additive effects combined with additive and multiplicative noise. Here a family of linear state space models with multiplicative noise and additive noise/forcing is considered with system equations formulated in the domain of the state variable, and a proposed Kalman filter for this model class is derived using the projection theorem. An obvious distribution assumption in the case of multiplicative noise is the log-*normal distribution, and an approximate likelihood function for such observations is suggested and applied for parameter estimation. The approach is exemplified by numerical simulation studies and the results are compared to the widely used maximum likelihood estimation under a local Gaussian assumption. The simulations demonstrate that it is advantageous to use the proposed method for estimation when the multiplicative noise shows substantial variation.*

Key Words: Multiplicative noise, Kalman filtering, maximum likelihood estimation, stochastic state space models, log-normal distribution

1 Introduction

In the analysis of dynamical systems it is often assumed that noise is additive and Gaussian. However, many real-life systems have scale-dependent noise

[1]DTU Informatics, Richard Pedersens Plads, Technical University of Denmark Building 321, DK-2800 Lyngby, Denmark.

[2]National Environmental Research Institute, University of Aarhus, Fredriksborgvej 399, DK-4000 Roskilde, Denmark.

components; examples include many financial time-series, biological growth series, air pollution series, etc. In biological growth models for example, the variance scales with the number of individuals. Biological growth models and other models that include various concentrations and process rates are, in addition to the scale-dependent noise, restricted to being positive. Positive and scale-dependent noise terms require other distributional assumptions than the Gaussian. For example Limpert et. al. (2001) emphasize the special role of the log-normal distribution in a broad range of disciplines. The example given below also suggests that the log-normal distribution has a special role in connection with multiplicative noise models.

Øksendal (2003) proposed a simple population growth model

$$\frac{dN_t}{dt} = \beta_t N_t, \tag{1}$$

where N_t is the amount of biological material and β_t is the growth rate. If the growth rate is subject to noise, then a natural way to express this uncertainty is $\beta_t = r + \alpha W_t$, where W_t is Gaussian white noise. The Stochastic Differential Equation (SDE) formulation corresponding to (1) is

$$dN_t = rN_t dt + \alpha N_t dB_t, \tag{2}$$

where $\int dB_s = B_t$ is (Gaussian) white noise. The (Itô) solution to (2) is

$$N_t = N_0 e^{(r-\frac{1}{2}\alpha^2)t+\alpha B_t}, \tag{3}$$

or in a discrete time framework

$$N_{t+1} = N_t e^{(r-\frac{1}{2}\alpha^2)+\alpha B_1} \tag{4}$$

$$= N_t e^r e^{\alpha B_1 - \frac{1}{2}\alpha^2} \tag{5}$$

$$= N_t a \xi_{t+1}, \tag{6}$$

where a is constant and ξ_t is a log-normal distributed multiplicative noise with expectation 1 (and variance $(e^{\alpha^2} - 1)$). In this example it is straightforward to transform (6) into a random Gaussian process by taking the logarithm. If the process had been multidimensional, or the SDE slightly more complex, then the solution might not be so straightforward. If e.g. (6) contains an external forcing like

$$N_{t+1} = N_t a \xi_{t+1} + f_{t+1}, \tag{7}$$

then taking the logarithm does not result in a simple linear form, and other measures are needed to treat this in a filtering setting.

Further, when data is transformed to obtain models that span the entire real axis and have constant variances, some of the phenomenological modeling structure

of the dynamical system will become difficult to formulate within a state space framework, and with this perspective in mind it can be advantageous to keep data and models in their original or natural domain. A focal point in the analysis of dynamical systems with log-normal distributed random variables is that the theory is far less developed than for Gaussian distributed state variables. The goal of this paper is to present results and techniques which allow a modeling of log-normally distributed processes in their original domain.

Section 2 presents the state space formulation for a system with multiplicative noise in the system equation as well as in the observation equation. For this system the dynamics are formulated in discrete time with discrete time observations, and hence it is referred to as a discrete-discrete system. In Section 3 we present a multi-dimensional linearized filter for the state space formulation presented in Section 2. The derivation of the filter is carried out in a general setting, i.e. with no distributional assumptions on the multiplicative noise.

The Kalman filter (Kalman (1960)), has enjoyed enormous success for linear Gaussian models. One way to derive the Kalman filter is by means of the projection theorem, see e.g. Madsen (2008). For Gaussian linear models the solution is exact, while for nonlinear and non-Gaussian models the solution is only approximate. Previous studies have investigated state space models with multiplicative noise; Chow and Birkemeier (1990) and Zhang and Zhang (2007) used the projection theorem to derive optimal filtering algorithms for models with multiplicative noise in the observations equation only. Yang et. al. (2002) derived a robust filter for systems with multiplicative scalar stochastic noise as well as "deterministic parametric noise", where the filter is designed to minimize an upper bound for the state variance. Jimenez and Ozaki (2002) presented a continuous-discrete filter with multiplicative and additive Gaussian noise terms. The multiplicative noise was introduced as coefficient noise, i.e. some (or all) of the coefficients have multiplicative noise. We introduce the multiplicative noise in both system and observation equations, i.e. multiplicative stochastic innovation acts on the states/observations seperately. A more fundamental difference between our approach and the previous methods is that we do not use the Gaussian assumption on the noise terms and furthermore we propose a likelihood estimation procedure under the assumption of log-normal multiplicative noise.

In this paper the projection theorem is used to derive the discrete-discrete filter for the system, the derivation of the filter for the system is presented in Section 3. Many real-life processes are positive by nature, some examples are given in Overgaard et. al. (2005) (pharmacokinetics), Vio et. al. (2005) (astronomy) and Guttal & Jayaprakash (2007) (ecology). Further, processes that involve concentrations will in general be restricted to being positive. It is therefore important to ensure that the state estimate of the system remains positive under

the projection approximations. For non-Gaussian noise the filter will only be approximate, but it is shown to work well with the log-normal distribution as presented in an example in Section 3.4.

In practice, parameters in models like Equation (6) are not known and must be estimated. In the non-linear and non-Gaussian case, the likelihood is often derived using Gaussian approximation to the one-step transition probabilities; continuous time examples of this are Kristensen et. al. (2004) and Nielsen et. al. (2000). We will derive an approximate log-likelihood procedure for positive valued systems with log-normal distributed multiplicative noise in the states and in the observations in Section 4. The estimation method is exemplified with simulation studies.

2 The State Space Formulation

A multivariate generalization of (6) is

$$\boldsymbol{X}_t = \boldsymbol{\Xi}_t \boldsymbol{A}_t \boldsymbol{X}_{t-1} + \boldsymbol{B}_t \boldsymbol{U}_{t-1}, \tag{8}$$

where $\boldsymbol{X}_t \in \mathbb{R}^n$ is the unobserved true state of the system. The input or the driver of the system, $\boldsymbol{U}_t \in \mathbb{R}^{n_u}$, can be stochastic or deterministic, but it is assumed to be independent of other variables in the system. If $\boldsymbol{U}_t$ is stochastic, then $\boldsymbol{U}_t$ should be given by its mean $\boldsymbol{\mu}_t^{\boldsymbol{U}}$ and variance $\boldsymbol{\Sigma}_t^{\boldsymbol{U}}$. Assume that the system is observed under multiplicative uncertainty

$$\boldsymbol{Y}_t = \boldsymbol{\Lambda}_t \boldsymbol{C}_t \boldsymbol{X}_t, \tag{9}$$

where $\boldsymbol{Y}_t \in \mathbb{R}^m$ is the observation of the system. In (8) and (9) it is assumed that the coefficient matrices $\boldsymbol{A}_t \in \mathbb{R}^{n \times n}$, $\boldsymbol{B}_t \in \mathbb{R}^{n \times n_u}$ and $\boldsymbol{C}_t \in \mathbb{R}^{m \times n}$ contain known parametric relationships between states, inputs and observations. $\boldsymbol{\Xi}_t \in \mathbb{R}^{n \times n}$ and $\boldsymbol{\Lambda}_t \in \mathbb{R}^{m \times m}$ are multiplicative noise terms, which are diagonal matrices with diagonal elements given as random vectors $\boldsymbol{\xi}_t \in \mathbb{R}^n$ ($\boldsymbol{\xi}_t = \text{diag}(\boldsymbol{\Xi}_t)$) and $\boldsymbol{\lambda}_t \in \mathbb{R}^m$ ($\boldsymbol{\lambda}_t = \text{diag}(\boldsymbol{\Lambda}_t)$). An equivalent formulation of (8)-(9) is therefore,

$$\boldsymbol{X}_t = \boldsymbol{\xi}_t \odot (\boldsymbol{A}_t \boldsymbol{X}_{t-1}) + \boldsymbol{B}_t \boldsymbol{U}_{t-1} \tag{10}$$

$$\boldsymbol{Y}_t = \boldsymbol{\lambda}_t \odot (\boldsymbol{C}_t \boldsymbol{X}_t), \tag{11}$$

where $\odot$ denotes Hadanmard (entry-wise) multiplication, $\boldsymbol{\xi}_t$ and $\boldsymbol{\lambda}_t$ have covariance matrices $\boldsymbol{\Sigma}_t^{\boldsymbol{\xi}}$ and $\boldsymbol{\Sigma}_t^{\boldsymbol{\lambda}}$, respectively, and expectation equal to unity. Furthermore, $\boldsymbol{\lambda}_t$ and $\boldsymbol{\xi}_t$ are assumed to be mutually independent white noise sequences. The formulation (10)-(11) emphasizes that the multiplicative noise acts on the states/observations, rather than on the coefficient matrices.

Many real-life processes are, as pointed out in the introduction, restricted to being positive. If $P\{\boldsymbol{\xi}_t < \mathbf{0}\} = P\{\boldsymbol{U}_t < \mathbf{0}\} = P\{\boldsymbol{\lambda}_t < \mathbf{0}\} = \mathbf{0}$, and $\boldsymbol{A}_t \geq \mathbf{0}$, $\boldsymbol{B}_t \geq \mathbf{0}$, $\boldsymbol{C}_t \geq \mathbf{0}$ then $\boldsymbol{X}_t \geq \mathbf{0}$ and $\boldsymbol{Y}_t \geq \mathbf{0}$ $\forall t$, provided that $\boldsymbol{X}_0 \geq \mathbf{0}$. With an appropriate choice of variables and parameters it is therefore possible to model processes in the original or natural scale rather than on a log scale. These requirements might seem quite restrictive, since it means that a positive change in one state can only cause a positive change in other states. However, if the family of models is extended to include non-linear models, then other correlation structures that allow negative interactions can be included and the filter can be applied in an Extended Kalman Filter (EKF) set up (see Jazwinski (1970)). An example of an ecological model with non-linear parameters which gives rise to such interactions is given in Møller et. al. (2009).

The formulation in (8)-(9) implies that the variance of the observations is zero when the state is zero. From a numerical point of view this constitutes a problem for the state updating procedure, and we will therefore use the following modified version of the observation equation

$$\boldsymbol{Y}_t = \boldsymbol{\Lambda}_t \boldsymbol{C}_t \boldsymbol{X}_t + \boldsymbol{\epsilon}_t, \tag{12}$$

where $\boldsymbol{\epsilon}_t \in \mathbb{R}^m$ is a random vector with mean $\boldsymbol{\mu}_t^{\boldsymbol{\epsilon}}$ and variance $\boldsymbol{\Sigma}_{\boldsymbol{t}}^{\boldsymbol{\epsilon}}$. If the process and the observations are positive, then $\boldsymbol{\epsilon}_t$ should be non-negative. If $\boldsymbol{\epsilon}_t$ is introduced for numerical reasons only, then the mean and variance should be small. Although $\boldsymbol{\epsilon}_t$ is imposed for numerical reasons, it will often make good sense in practice. The construction corresponds to a situation where the variance of the observations scales with the size of the population when the population is large. When the population approaches zero the observations are still subject to a small additive observation error.

3 The Filter Equations

The aim of this section is to derive expressions for the recursive filter equations related to the state space model as specified in (8)-(9). These consist of the reconstruction

$$\hat{\boldsymbol{X}}_{t|t} = E[\boldsymbol{X}_t|\mathcal{Y}_t] \tag{13}$$

$$\boldsymbol{\Sigma}_{t|t}^{xx} = Var[\boldsymbol{X}_t|\mathcal{Y}_t], \tag{14}$$

and the predictions

$$\hat{\boldsymbol{X}}_{t+1|t} = E[\boldsymbol{X}_{t+1}|\mathcal{Y}_t] \tag{15}$$
$$\boldsymbol{\Sigma}^{xx}_{t+1|t} = Var[\boldsymbol{X}_{t+1}|\mathcal{Y}_t] \tag{16}$$
$$\hat{\boldsymbol{Y}}_{t+1|t} = E[\boldsymbol{Y}_{t+1}|\mathcal{Y}_t] \tag{17}$$
$$\boldsymbol{\Sigma}^{yy}_{t+1|t} = Var[\boldsymbol{Y}_{t+1}|\mathcal{Y}_t] \tag{18}$$
$$\boldsymbol{\Sigma}^{xy}_{t+1|t} = C[\boldsymbol{X}_{t+1}, \boldsymbol{Y}_{t+1}|\mathcal{Y}_t], \tag{19}$$

where $\mathcal{Y}_t$ denotes the information at time t, i.e. the observations and the input. The results for linear systems with additive and Gaussian noise are well-known and provided by the Kalman filter recursions, see e.g. Madsen (2008). The system considered here, however, is a system with multiplicative non-Gaussian noise. In the following the resulting filter will be applied for maximum likelihood estimation of the parameters in equation (8)-(9). The estimation framework is described in Section 4 under the assumption that the multiplicative noise terms and input are log-normal distributed.

3.1 The Reconstruction

The linear reconstruction of the states is given by the linear projection (e.g. Madsen (2008) p. 24) and is approximated by

$$\hat{\boldsymbol{X}}_{t|t} \simeq \hat{\boldsymbol{X}}_{t|t-1} + \boldsymbol{\Sigma}^{xy}_{t|t-1} \left(\boldsymbol{\Sigma}^{yy}_{t|t-1}\right)^{-1} (\boldsymbol{Y}_t - \hat{\boldsymbol{Y}}_{t|t-1}) \tag{20}$$

Equation (20) is exact in the linear Gaussian case. The assumption in the following is therefore that the process can be approximated locally by the linearized process. For notational convenience we define $\hat{\boldsymbol{X}}_{t|t}$ as the approximation in (20), and with this we get

$$\hat{\boldsymbol{X}}_{t|t} = \hat{\boldsymbol{X}}_{t|t-1} + \boldsymbol{\Sigma}^{xy}_{t|t-1}(\boldsymbol{\Sigma}^{yy}_{t|t-1})^{-1}(\boldsymbol{Y}_t - \boldsymbol{C}_t\hat{\boldsymbol{X}}_{t|t-1} - \boldsymbol{\mu}^{\boldsymbol{\epsilon}}_t), \tag{21}$$

where we have used that $\boldsymbol{\Lambda}_t$ is independent of $\boldsymbol{X}_t$ and $E[\boldsymbol{\Lambda}_t] = \boldsymbol{I}$. Equation (21) can produce negative reconstructions even when the setup is such that the states are positive with probability one. In this case we simply set the i'th state equal to zero if the reconstruction is less than zero, i.e.

$$(\hat{\boldsymbol{X}}_{t|t})_i = \max\{0, (\hat{\boldsymbol{X}}_{t|t-1} + \boldsymbol{\Sigma}^{xy}_{t|t-1}(\boldsymbol{\Sigma}^{yy}_{t|t-1})^{-1}(\boldsymbol{Y}_t - \boldsymbol{C}_t\hat{\boldsymbol{X}}_{t|t-1} - \boldsymbol{\mu}^{\boldsymbol{\epsilon}}_t))_i\}. \tag{22}$$

The variance of the reconstruction of the state is approximated by

$$\boldsymbol{\Sigma}^{xx}_{t|t} \simeq \boldsymbol{\Sigma}^{xx}_{t|t-1} - \boldsymbol{\Sigma}^{xy}_{t|t-1}(\boldsymbol{\Sigma}^{yy}_{t|t-1})^{-1}(\boldsymbol{\Sigma}^{xy}_{t|t-1})^T. \tag{23}$$

Again the assumption is that the process can be approximated well by the linearised process. Thus the reconstruction is calculated using the predictions derived below. The derivations so far are the same as for the linear Gaussian case and $\boldsymbol{K}_t = \boldsymbol{\Sigma}^{xy}_{t|t-1}(\boldsymbol{\Sigma}^{yy}_{t|t-1})^{-1}$ is recognized as the Kalman gain. In the linear Gaussian case this will converge to the steady-state Kalman recursions, and if the coefficients in addition are constant in time, then the Kalman gain will converge to a constant matrix which is not influenced by random fluctuation - neither in the state nor in the observations. In the multiplicative case the decoupling of state and variance does not exist, as the state of the system will have an impact on the state variance.

3.2 The Predictions

The prediction of the state $\hat{\boldsymbol{X}}_{t+1|t}$ is

$$\begin{aligned}\hat{\boldsymbol{X}}_{t+1|t} &= E[\boldsymbol{\Xi}_{t+1}\boldsymbol{A}_{t+1}\boldsymbol{X}_t + \boldsymbol{B}_{t+1}\boldsymbol{U}_t|\mathcal{Y}_t] &(24)\\ &= E[\boldsymbol{\Xi}_{t+1}|\mathcal{Y}_t]E[\boldsymbol{A}_t\boldsymbol{X}_t|\mathcal{Y}_t] + \boldsymbol{B}_{t+1}E[\boldsymbol{U}_t|\mathcal{Y}_t] &(25)\\ &= \boldsymbol{A}_t\hat{\boldsymbol{X}}_{t|t} + \boldsymbol{B}_{t+1}\boldsymbol{\mu}^{\boldsymbol{u}}_t. &(26)\end{aligned}$$

The prediction of the state covariance is

$$\begin{aligned}\boldsymbol{\Sigma}^{xx}_{t+1|t} =& Var[\boldsymbol{\Xi}_{t+1}\boldsymbol{A}_{t+1}\boldsymbol{X}_t + \boldsymbol{B}_{t+1}\boldsymbol{U}_t|\mathcal{Y}_t] &(27)\\ =& E[(\boldsymbol{\Xi}_{t+1}\boldsymbol{A}_{t+1}\boldsymbol{X}_t - \boldsymbol{A}_{t+1}\hat{\boldsymbol{X}}_{t|t})(\boldsymbol{\Xi}_{t+1}\boldsymbol{A}_{t+1}\boldsymbol{X}_t - \boldsymbol{A}_{t+1}\hat{\boldsymbol{X}}_{t|t})^T|\mathcal{Y}_t] +\\ & \boldsymbol{B}_{t+1}\boldsymbol{\Sigma}^{\boldsymbol{u}}_t\boldsymbol{B}^T_{t+1} &(28)\\ =& E[\boldsymbol{\Xi}_{t+1}\boldsymbol{A}_{t+1}\boldsymbol{X}_t\boldsymbol{X}^T_t\boldsymbol{A}^T_{t+1}\boldsymbol{\Xi}^T_{t+1}|\mathcal{Y}_t] - \boldsymbol{A}_{t+1}\hat{\boldsymbol{X}}_{t|t}\hat{\boldsymbol{X}}^T_{t|t}\boldsymbol{A}^T_{t+1} +\\ & \boldsymbol{B}_{t+1}\boldsymbol{\Sigma}^{\boldsymbol{u}}_t\boldsymbol{B}^T_{t+1}. &(29)\end{aligned}$$

The independence between $\boldsymbol{\Xi}_{t+1}$ and $\boldsymbol{X}_{t+1}$ is used to obtain (28) and (29). In order to derive an expression for the expectation in (29) we look at the individual elements of the matrix, i.e.

$$\begin{aligned}&(E[\boldsymbol{\Xi}_{t+1}\boldsymbol{A}_{t+1}\boldsymbol{X}_t\boldsymbol{X}^T_t\boldsymbol{A}^T_{t+1}\boldsymbol{\Xi}^T_{t+1}|\mathcal{Y}_t])_{i,j} &(30)\\ &= E[\xi^{t+1}_i(\boldsymbol{A}_{t+1}\boldsymbol{X}_t\boldsymbol{X}^T_t\boldsymbol{A}^T_{t+1})_{i,j}\xi^{t+1}_j|\mathcal{Y}_t] &(31)\\ &= E[\xi^{t+1}_i\xi^{t+1}_j|\mathcal{Y}_t]E[(\boldsymbol{A}_{t+1}\boldsymbol{X}_t\boldsymbol{X}^T_t\boldsymbol{A}^T_{t+1})_{i,j}|\mathcal{Y}_t]. &(32)\end{aligned}$$

The first expectation in (32) is

$$\begin{aligned}E[\xi^{t+1}_i\xi^{t+1}_j|\mathcal{Y}_t] &= Cov[\xi^{t+1}_i, \xi^{t+1}_j] + E[\xi^{t+1}_i]E[\xi^{t+1}_j] &(33)\\ &= \sigma^{\boldsymbol{\xi}_{t+1}}_{i,j} + 1. &(34)\end{aligned}$$

This can be written in matrix notation as $E[\boldsymbol{\xi}_{t+1}\boldsymbol{\xi}_{t+1}^T] = \boldsymbol{\Sigma}_{t+1}^{\boldsymbol{\xi}} + \boldsymbol{e}\boldsymbol{e}^T$, where $\boldsymbol{e}$ is a vector of ones, the matrix $\boldsymbol{e}\boldsymbol{e}^T$ works as the identity matrix w.r.t. element-wise multiplication, i.e. $(\boldsymbol{e}\boldsymbol{e}^T) \odot \boldsymbol{F} = \boldsymbol{F}$, $\forall \boldsymbol{F} \in \mathbb{R}^{n\times n}$. The second expectation in (32) can be written as

$$\begin{aligned} E[\boldsymbol{A}_{t+1}\boldsymbol{X}_t\boldsymbol{X}_t^T\boldsymbol{A}_{t+1}^T|\mathcal{Y}_t] &= \boldsymbol{A}_{t+1}(Var[\boldsymbol{X}_t|\mathcal{Y}_t] + E[\boldsymbol{X}_t|\mathcal{Y}_t]E[\boldsymbol{X}_t^T|\mathcal{Y}_t])\boldsymbol{A}_{t+1}^T \\ &= \boldsymbol{A}_{t+1}(\boldsymbol{\Sigma}_{t|t}^{xx} + \hat{\boldsymbol{X}}_{t|t}\hat{\boldsymbol{X}}_{t|t}^T)\boldsymbol{A}_{t+1}^T. \end{aligned} \tag{35}$$

Denote the second order non-central moment by

$$\hat{\boldsymbol{P}}_{t+k|t}^{xx} = E[\boldsymbol{X}_{t+k}\boldsymbol{X}_{t+k}^T|\mathcal{Y}_t] = \boldsymbol{\Sigma}_{t+k|t}^{xx} + \hat{\boldsymbol{X}}_{t+k|t}\hat{\boldsymbol{X}}_{t+k|t}^T. \tag{36}$$

Combining (32), (34) and (35) gives

$$E[\boldsymbol{\Xi}_{t+1}\boldsymbol{A}_{t+1}\boldsymbol{X}_t\boldsymbol{X}_t^T\boldsymbol{A}_{t+1}^T\boldsymbol{\Xi}_{t+1}^T|\mathcal{Y}_t] = (\boldsymbol{\Sigma}_{t+1}^{\boldsymbol{\xi}} + \boldsymbol{e}\boldsymbol{e}^T) \odot (\boldsymbol{A}_{t+1}\hat{\boldsymbol{P}}_{t|t}^{xx}\boldsymbol{A}_{t+1}^T) \tag{37}$$

$$= \boldsymbol{\Sigma}_{t+1}^{\boldsymbol{\xi}} \odot (\boldsymbol{A}_{t+1}\hat{\boldsymbol{P}}_{t|t}^{xx}\boldsymbol{A}_{t+1}^T) + \boldsymbol{A}_{t+1}\hat{\boldsymbol{P}}_{t|t}\boldsymbol{A}_{t+1}^T. \tag{38}$$

Inserting (38) in (29) gives the variance of the state prediction

$$\boldsymbol{\Sigma}_{t+1|t}^{xx} = \boldsymbol{A}_{t+1}\boldsymbol{\Sigma}_{t|t}^{xx}\boldsymbol{A}_{t+1}^T + \boldsymbol{\Sigma}_{t+1}^{\boldsymbol{\xi}} \odot (\boldsymbol{A}_{t+1}\hat{\boldsymbol{P}}_{t|t}^{xx}\boldsymbol{A}_{t+1}^T) + \boldsymbol{B}_{t+1}\boldsymbol{\Sigma}_{t+1}^{\boldsymbol{u}}\boldsymbol{B}_{t+1}^T. \tag{39}$$

From (39) it is seen that the state has a direct impact on the variance through $\hat{\boldsymbol{P}}_{t|t}^{xx}$, resulting in the scale-dependent variance prediction. The prediction of the new observation is

$$\hat{\boldsymbol{Y}}_{t+1|t} = \boldsymbol{C}_{t+1}\hat{\boldsymbol{X}}_{t+1|t} + \boldsymbol{\mu}_{t+1}^{\boldsymbol{\epsilon}}. \tag{40}$$

The covariance of the predicted observation $\boldsymbol{\Sigma}_{t+1|t}^{yy}$ is calculated by the same technique as the predicted state variance and the result is

$$\boldsymbol{\Sigma}_{t+1|t}^{yy} = Var[\boldsymbol{\Lambda}_{t+1}\boldsymbol{C}_{t+1}\boldsymbol{X}_{t+1} + \boldsymbol{\epsilon}_{t+1}|\mathcal{Y}_t] \tag{41}$$

$$= \boldsymbol{\Sigma}_{t+1}^{\boldsymbol{\epsilon}} - \boldsymbol{C}_{t+1}\hat{\boldsymbol{X}}_{t+1|t}\hat{\boldsymbol{X}}_{t+1|t}^T\boldsymbol{C}_{t+1}^T + E[\boldsymbol{\Lambda}_{t+1}\boldsymbol{C}_{t+1}\boldsymbol{X}_{t+1}\boldsymbol{X}_{t+1}^T\boldsymbol{C}_{t+1}^T\boldsymbol{\Lambda}_{t+1}^T|\mathcal{Y}_t] \tag{42}$$

$$= \boldsymbol{\Sigma}_{t+1}^{\boldsymbol{\epsilon}} + \boldsymbol{C}_{t+1}\boldsymbol{\Sigma}_{t+1|t}^{xx}\boldsymbol{C}_{t+1}^T + \boldsymbol{\Sigma}_{t+1}^{\boldsymbol{\lambda}} \odot (\boldsymbol{C}_{t+1}\hat{\boldsymbol{P}}_{t+1|t}^{xx}\boldsymbol{C}_{t+1}^T). \tag{43}$$

The covariance ($\boldsymbol{\Sigma}_{t|t-1}^{xy}$) between the predicted observation and the predicted state is calculated by

$$\boldsymbol{\Sigma}_{t|t-1}^{xy} = C[\boldsymbol{X}_t, \boldsymbol{\Lambda}_t\boldsymbol{C}_t\boldsymbol{X}_t + \boldsymbol{\epsilon}_t|\mathcal{Y}_{t-1}] \tag{44}$$

$$= E[(\boldsymbol{X}_t - E[\boldsymbol{X}_t|\mathcal{Y}_{t-1}])(\boldsymbol{\Lambda}_t\boldsymbol{C}_t\boldsymbol{X}_t - E[\boldsymbol{\Lambda}_t\boldsymbol{C}_t\boldsymbol{X}_t|\mathcal{Y}_{t-1}])^T|\mathcal{Y}_{t-1}] \tag{45}$$

$$= \boldsymbol{\Sigma}_{t|t-1}^{xx}\boldsymbol{C}_t^T, \tag{46}$$

where the independence between $\boldsymbol{\lambda}_t$, $\boldsymbol{\epsilon}_t$ and $\boldsymbol{X}_t$ is used to obtain (45).

3.3 Summary

In summary, we have the reconstruction

$$\hat{\boldsymbol{X}}_{t|t} = \hat{\boldsymbol{X}}_{t|t-1} + \boldsymbol{K}_t(\boldsymbol{Y}_t - \hat{\boldsymbol{Y}}_{t|t-1}) \tag{47}$$

$$\boldsymbol{\Sigma}^{xx}_{t|t} = \boldsymbol{\Sigma}^{xx}_{t|t-1} - \boldsymbol{K}_t(\boldsymbol{\Sigma}^{xy}_{t|t-1})^T. \tag{48}$$

and the predictions

$$\hat{\boldsymbol{X}}_{t+1|t} = \boldsymbol{A}_{t+1}\hat{\boldsymbol{X}}_{t|t} + \boldsymbol{B}_{t+1}\boldsymbol{\mu}^{\boldsymbol{u}}_t \tag{49}$$

$$\boldsymbol{\Sigma}^{xx}_{t+1|t} = \boldsymbol{A}_{t+1}\boldsymbol{\Sigma}^{xx}_{t|t}\boldsymbol{A}^T_{t+1} + \boldsymbol{\Sigma}^{\boldsymbol{\xi}}_{t+1} \odot (\boldsymbol{A}_{t+1}\boldsymbol{P}^{xx}_{t|t}\boldsymbol{A}^T_{t+1}) + \boldsymbol{B}_{t+1}\boldsymbol{\Sigma}^{\boldsymbol{u}}_t\boldsymbol{B}^T_{t+1} \tag{50}$$

$$\hat{\boldsymbol{Y}}_{t+1|t} = \boldsymbol{C}_{t+1}\hat{\boldsymbol{X}}_{t+1|t} + \boldsymbol{\mu}^{\boldsymbol{\epsilon}}_t \tag{51}$$

$$\boldsymbol{\Sigma}^{yy}_{t+1|t} = \boldsymbol{C}_{t+1}\boldsymbol{\Sigma}^{xx}_{t+1|t}\boldsymbol{C}^T_{t+1} + \boldsymbol{\Sigma}^{\boldsymbol{\lambda}}_{t+1} \odot (\boldsymbol{C}_{t+1}\boldsymbol{P}^{xx}_{t+1|t}\boldsymbol{C}^T_{t+1}) + \boldsymbol{\Sigma}^{\boldsymbol{\epsilon}}_{t+1} \tag{52}$$

$$\boldsymbol{\Sigma}^{xy}_{t+1|t} = \boldsymbol{\Sigma}^{xx}_{t+1|t}\boldsymbol{C}^T_{t+1}, \tag{53}$$

with

$$\boldsymbol{K}_t = \boldsymbol{\Sigma}^{xx}_{t|t-1}\boldsymbol{C}^T_t(\boldsymbol{\Sigma}^{yy}_{t|t-1})^{-1} \tag{54}$$

$$\boldsymbol{P}^{xx}_{t+k|t} = \boldsymbol{\Sigma}^{xx}_{t+k|t} + \hat{\boldsymbol{X}}_{t+k|t}\hat{\boldsymbol{X}}^T_{t+k|t}, \tag{55}$$

and given initial conditions

$$\hat{\boldsymbol{X}}_{1|0} = \boldsymbol{\mu}_0, \quad \hat{\boldsymbol{\Sigma}}^{xx}_{1|0} = \boldsymbol{V}_0. \tag{56}$$

If the dynamics are such that the process is positive for all t then (47) is replaced by

$$(\hat{\boldsymbol{X}}_{t|t})_i = \max\{0, (\hat{\boldsymbol{X}}_{t|t-1} + \boldsymbol{\Sigma}^{xy}_{t|t-1}(\boldsymbol{\Sigma}^{yy}_{t|t-1})^{-1}(\boldsymbol{Y}_t - \hat{\boldsymbol{Y}}_{t|t-1}))_i\}. \tag{57}$$

These recursive equations will form the basis for the likelihood estimation in Section 4. Section 3.4 presents an application of the filter with simulated data.

3.4 Example

The following state space model is considered

$$\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} \xi_{1,t} & 0 \\ 0 & \xi_{2,t} \end{bmatrix} \begin{bmatrix} 0.5 & 0 \\ 0.1 & 0.7 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} U_t, \tag{58}$$

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = \begin{bmatrix} \lambda_{1,t} & 0 \\ 0 & \lambda_{2,t} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}, \tag{59}$$

with $\boldsymbol{\xi}_t$ and $\boldsymbol{\lambda}_t$ log-normal distributed random variables with covariances

$$\boldsymbol{\Sigma}_t^{\boldsymbol{\xi}} = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.4 \end{bmatrix}, \quad \boldsymbol{\Sigma}_t^{\boldsymbol{\lambda}} = \begin{bmatrix} 0.6 & 0 \\ 0 & 0.6 \end{bmatrix}, \tag{60}$$

respectively and expectation equal to the identity matrix. $\epsilon_{1,t}$ and $\epsilon_{2,t}$ are independent iid sequences of exponentially distributed random variables with mean 0.01. The initial values are set at

$$\hat{\boldsymbol{X}}_{1|0} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{\Sigma}_{1|0}^{xx} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{61}$$

U_t is a sequence of random variables of the form

$$U_t = \mu_t^u r_t, \tag{62}$$

where r_t is a log-normal distributed random variable with mean 1 and variance 0.1. μ_t^u is a deterministic sequence which spans several orders of magnitude (see simulations in Figure 1). Figure 1 illustrates that the filter captures the dynamics of the simulated process when the states are large, and Figure 1 (right column) illustrates how the filter works for small values of input and states. Note also that the variation of the process around the mean is stable over the entire range of values, when considered in the log-domain.

Confidence regions are constructed under the assumption that sums of log-normal distributed random variables follow another log-normal distribution. This assumption is supported by Fenton (1960), who points out that the distribution of a sum of log-normal variables is well approximated by a distribution function for a log-normal random variable with matching (of the sums) first and second moment for the middle range of the random variable, while higher order moments should be matched for higher values of the random variable. In the following we assume that the results from Fenton (1960) can be generalized to the multivariate case. This assumption is not proven or pursued any further here, but it seems to be valid for the simulations performed here. The filter provides the first and second moments and these are therefore used for the approximation here. The covariance matrix obtained by the filter is positive definite, but this restriction is not sufficient for a covariance matrix for a multidimensional log-normal random variable. We will refer to covariance matrices which meet the requirements for log-normal covariance matrices as admissible. To ensure admissibility we transform the covariance matrices for the state reconstruction and the observation predictions by (88), on page 104. The estimated confidence region is assessed in the normal domain and in our example 95.7 % of the simulated true state and 95.9 % of the observations were inside the confidence regions. That the fraction of observations and states inside the confidence regions are close to the required fractions supports the multivariate log-normal assumption.

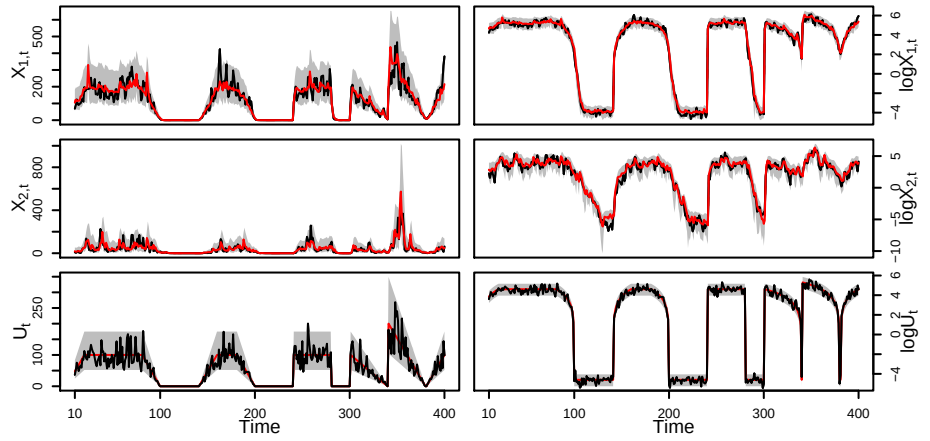

Figure 1: The true state $\boldsymbol{X}_t = (X_{1,t}, X_{2,t})$ of the system (black), the state reconstruction $\hat{\boldsymbol{X}}_{t|t}$ of the system (red), with 95% confidence intervals (grey area), forced by the stochastic input U_t (black), with mean μ_t^U (red) and 95% confidence interval for input (grey area). The first 10 datapoints are considered as transients and are therefore not shown.

The confidence regions cannot be visualized in plots like the ones presented in Figure 1. The confidence intervals in Figure 1 are therefore calculated as confidence intervals for the unconditional process, i.e. for a univariate random variable with expectation $(\hat{\boldsymbol{X}}_{t|t})_i$ and variance $(\boldsymbol{\Sigma}^{xx}_{t|t})_{i,i}$.

The dynamics of the simulated process are well reproduced by the reconstruction, even during strong, abrupt perturbations of U_t. It is seen that the variance of the state reconstruction is almost constant in the log-domain. It is important to stress that even when we consider the estimated state in the log-domain, the filtering has actually been carried out in the original domain, and the formulation of the dynamic interaction was also done in the original domain. This is important since it is often easier to do the reasoning in the natural domain than in the log-domain.

For linear Gaussian processes with constant parameters, the variance of the noise and the input is often assumed to be constant and for the classical Kalman filter case the variance of the state estimates will then converge to some constant matrix. This is not the case for models with multiplicative noise, where the variance is a function of the mean and this scale dependency of the noise is an important property of the process which should be emphasized.

While confidence intervals are symmetric around the mean value for Gaussian processes, this is not the case for log-normal random variables. Limpert et. al. (2001) point out that while confidence intervals for Gaussian random variables are characterized by $\mu \pm a \cdot \sigma$ (where $a = 1$ leads to 68.3% confidence intervals and $a = 2$ leads to 95.5% confidence intervals), then log-normal random variables are characterized by $\mu^{*\times}/\sigma^{*a}$. Hence a 68.3% confidence interval is $[\mu^*/\sigma^*, \mu^* \cdot \sigma^*]$ and a 95.5% confidence interval is $[\mu^*/\sigma^{*2}, \mu^* \cdot \sigma^{*2}]$, where $\mu^* = e^{\mu}$ is the median for the log-normal distribution and $\sigma^* = e^{\sigma}$ is the multiplicative standard error.

Following Limpert et. al. (2001) we will denote the filter version of the multiplicative standard deviation (estimate of σ^*) by s. In the notation of the filter equation, the result for the i'th state is

$$s_{i,t}^{xx} = \exp\left(\sqrt{\log\left(\frac{(\mathbf{\Sigma}_{t|t}^{xx})_{ii}}{(\hat{\boldsymbol{X}}_{t|t})_i^2} + 1\right)}\right), \tag{63}$$

where $\sqrt{\log\left((\mathbf{\Sigma}_{t|t}^{xx})_{ii}/(\hat{\boldsymbol{X}}_{t|t})_i^2 + 1\right)}$ is the standard deviation for the corresponding normal distribution (this expression can be derived by considering equations (68)-(73) given in the beginning of Section 4).

Comparing the standard deviation ($\sigma_i^{xx} = \sqrt{(\mathbf{\Sigma}_{t|t}^{xx})_{ii}}$) with the multiplicative standard deviation ($s_{i,t}^{xx}$) (Figure 2), it is seen that while the standard deviation varies with the mean value, there is much smaller variation in the multiplicative standard deviation. According to Limpert et. al. (2001), s can be seen as a measure of how far the distribution is from a symmetric distribution. Limpert et. al. (2001) report empirical multiplicative standard deviations in the range from 1.16 to 33.15 covering a broad range of disciplines. The relative constant multiplicative standard deviation observed in Figure 2 is similarly reflected in the relatively constant confidence intervals in Figure 1 (right column).

The correlation between $X_{1,t}$ and $X_{2,t}$ is seen to vary; occasionally abruptly. These changes are explained by the sudden changes in the input, which also lead to strong changes in the variance of $X_{1,t}$. This can be further elaborated by examining how changes in $\mathbf{\Sigma}_t^U$ effect the state variance (consider equation (50) and (48), with the parameters given in this example). If changes in the loading had been more gradual, then these abrupt changes in the correlations would not have occurred, since the covariance would then be able to adjust gradually to the changes in input variance.

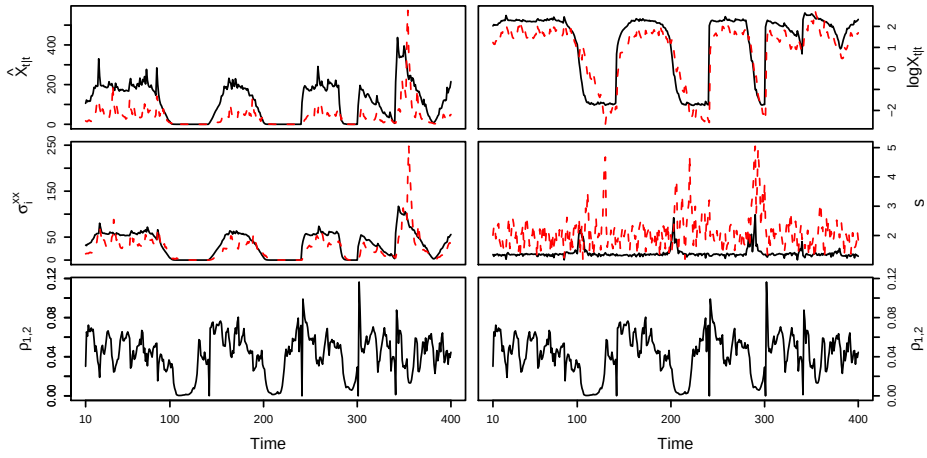

Figure 2: Left panel: the states (State 1, black and State 2, red), the standard deviation of the states and the correlation between the states on the original scale. Right panel: the logarithm of the states, the "multiplicative" standard deviation and the correlation.

4 Likelihood Estimation

Likelihood estimation in classical Gaussian state space models is well-known and based on the fact that sums of Gaussian random variables are also Gaussian random variables. An example on the derivation of the likelihood function for linear Gaussian state space models can be found in Chapter 10 of Madsen (2008). For non-linear and non-Gaussian models it is often assumed that the conditional densities of the process are local Gaussian, see e.g. Kristensen et. al. (2004) and Overgaard et. al. (2005). The result of this assumption is the log-likelihood function

$$\log L_G(\boldsymbol{\theta}; \mathcal{Y}_N) = -\frac{1}{2}\sum_{i=1}^{N}\left(\log(\det \boldsymbol{\Sigma}^{yy}_{i|i-1}) + \tilde{\boldsymbol{Y}}^T_{i|i-1}(\boldsymbol{\Sigma}^{yy}_{i|i-1})^{-1}\tilde{\boldsymbol{Y}}_{i|i-1}\right) + \text{constant}, \tag{64}$$

where $\tilde{\boldsymbol{Y}}_{i|i-1} = \boldsymbol{Y}_i - \hat{\boldsymbol{Y}}_{i|i-1}$. Maximizing (64) w.r.t. the parameters $\boldsymbol{\theta}$ gives the (Gaussian) ML estimate

$$\hat{\boldsymbol{\theta}}_G = \arg\left\{\max_{\theta} \log L_G(\boldsymbol{\theta}; \mathcal{Y}_N)\right\}. \tag{65}$$

In this section we will derive an approximate likelihood function under the assumption that the noise terms are log-normal distributed.

In the following it is assumed, see also Section 2, that $\boldsymbol{A}_t \geq \boldsymbol{0}$, $\boldsymbol{B}_t \geq \boldsymbol{0}$, $\boldsymbol{C}_t \geq \boldsymbol{0}$ and that $\boldsymbol{\xi}_t$, $\boldsymbol{\lambda}_t$ and $\boldsymbol{U}_t$ follow independent (multivariate) log-normal distributions ($P\{\boldsymbol{\xi}_t < \boldsymbol{0}\} = P\{\boldsymbol{\lambda}_t < \boldsymbol{0}\} = P\{\boldsymbol{U}_t < \boldsymbol{0}\} = \boldsymbol{0}$). For the system considered here, the one step prediction of the states is a sum of log-normal distributed random variables. The distribution of such a sum cannot, in general, be written in closed form. Using the results from Fenton (1960), as discussed in Section 3.4, it is assumed that the states and the observations can be approximated by log-normal random variables. In the derivations below we will ignore the contributions from $\boldsymbol{\epsilon}_t$, and the approximation can only be expected to be good when the variance of $\boldsymbol{\epsilon}_t$ is relatively small compared to the variance contributions from the states, see equation (52).

Using the assumption above the observations will be approximately log-normal distributed with expectation $\hat{\boldsymbol{Y}}_{t+1|t} = \boldsymbol{C}_t \hat{\boldsymbol{X}}_{t+1|t}$ and variance $\boldsymbol{\Sigma}^{yy}_{t+1|t}$. If $\boldsymbol{Y}_t$ follows a multivariate log-normal distribution then $\boldsymbol{Z}_t = \log(\boldsymbol{Y}_t)$ follows a multivariate normal distribution with expectation $\hat{\boldsymbol{Z}}_{t+1|t}$ and variance equal $\boldsymbol{\Sigma}^{zz}_{t+1|t}$. Setting $(\boldsymbol{\Sigma}^{zz}_{t+1|t})_{i,j} = \sigma^{zz}_{i,j}$ and $(\boldsymbol{\Sigma}^{yy}_{t+1|t})_{i,j} = \sigma^{yy}_{i,j}$ then (see Kotz et. al. (2000))

$$\hat{Y}_{i,t+1|t} = e^{\hat{Z}_{i,t+1|t} + \frac{1}{2}\sigma^{zz}_{i,i}} \Rightarrow \log\left(\hat{Y}_{i,t+1|t}\right) = \hat{Z}_{i,t+1|t} + \frac{1}{2}\sigma^{zz}_{i,i} \tag{66}$$

$$\sigma^{yy}_{i,j} = (e^{\sigma^{zz}_{i,j}} - 1) e^{\hat{Z}_{i,t+1|t} + \hat{Z}_{j,t+1|t} + \frac{1}{2}(\sigma^{zz}_{i,i} + \sigma^{zz}_{j,j})} \tag{67}$$

$$= (e^{\sigma^{zz}_{i,j}} - 1)\hat{Y}_{i,t+1|t}\hat{Y}_{j,t+1|t}, \tag{68}$$

rearranging (66) and (68) gives

$$\sigma^{zz}_{i,j} = \log\left(\frac{\sigma^{yy}_{i,j} + \hat{Y}_{i,t+1|t}\hat{Y}_{j,t+1|t}}{\hat{Y}_{i,t+1|t}\hat{Y}_{j,t+1|t}}\right) \tag{69}$$

$$= \log\left(\sigma^{yy}_{i,j} + \hat{Y}_{i,t+1|t}\hat{Y}_{j,t+1|t}\right) - \log\left(\hat{Y}_{i,t+1|t}\hat{Y}_{j,t+1|t}\right) \tag{70}$$

$$\hat{Z}_{i,t+1|t} = \log\left(\hat{Y}_{i,t+1|t}\right) - \frac{1}{2}\sigma^{zz}_{i,i}. \tag{71}$$

In matrix notation this is

$$\boldsymbol{\Sigma}^{zz}_{t+1|t} = \log\left(\boldsymbol{\Sigma}^{yy}_{t+1|t} + \hat{\boldsymbol{Y}}_{t+1|t}\hat{\boldsymbol{Y}}^{T}_{t+1|t}\right) - \log\left(\hat{\boldsymbol{Y}}_{t+1|t}\hat{\boldsymbol{Y}}^{T}_{t+1|t})\right) \tag{72}$$

$$\hat{\boldsymbol{Z}}_{t+1|t} = \log\left(\hat{\boldsymbol{Y}}_{t+1|t}\right) - \frac{1}{2}\mathrm{diag}\boldsymbol{\Sigma}^{zz}_{t+1|t}, \tag{73}$$

The variance of $\boldsymbol{Y}_{t+1}|\mathcal{Y}_t$ is calculated using a linear projection and it is guaranteed to be a variance matrix in the sense that it is positive definite and it can

be written as

$$\left(\boldsymbol{\Sigma}^{yy}_{t+1|t}\right)_{i,j} = \begin{cases} \sigma^{yy}_{ii} & \text{for} \quad i = j \\ \rho^{yy}_{ij}\sqrt{\sigma^{yy}_{ii}\sigma^{yy}_{jj}} & \text{for} \quad i \neq j \end{cases} \tag{74}$$

where $\rho^{yy}_{ij} \in (-1, 1)$. If $\boldsymbol{Y}_{t+1|t}$ is a multidimensional log-normal random variable, then only a subset of the correlations in $(-1, 1)$ are admissible. This subset is determined by the mean and the variance. We will seek a transformation of $\rho^{yy}_{i,j}$ such that the covariance matrix is guaranteed to be admissible. For this transformation and its practical implementation the implicit assumption is that the variances and the mean values are most precisely estimated and the correlation only has to be adjusted to ensure admissibility.

If $\boldsymbol{Z}_t$ is a Gaussian random variable, such that $\boldsymbol{Z}_t = \log(\boldsymbol{Y}_t)$, then

$$\rho^{yy}_{ij} = \frac{e^{\rho^{zz}_{ij}\sqrt{\sigma^{zz}_{ii}\sigma^{zz}_{jj}}} - 1}{\sqrt{(e^{\sigma^{zz}_{ii}} - 1)(e^{\sigma^{zz}_{jj}} - 1)}}. \tag{75}$$

Since $\rho^{zz}_{ij} \in (-1, 1)$ and ρ^{yy}_{ij} is a monotone increasing function of ρ^{zz}_{ij}, it is possible to calculate the maximum and minimum possible values of ρ^{yy}_{ij}. These will be

$$\rho^{yy}_{ij,max} = \frac{e^{\sqrt{\sigma^{zz}_{ii}\sigma^{z}_{jj}}} - 1}{\sqrt{(e^{\sigma^{zz}_{ii}} - 1)(e^{\sigma^{zz}_{jj}} - 1)}}, \quad \rho^{yy}_{ij,min} = \frac{e^{-\sqrt{\sigma^{zz}_{ii}\sigma^{z}_{jj}}} - 1}{\sqrt{(e^{\sigma^{zz}_{ii}} - 1)(e^{\sigma^{zz}_{jj}} - 1)}}. \tag{76}$$

We now seek a monotone increasing transformation $\tilde{\rho}^{yy}_{ij}(\rho^{yy}_{ij})$, s.t. $\tilde{\rho}^{yy}_{ij}(1) = \rho^{yy}_{ij,max}$, $\tilde{\rho}^{yy}_{i,j}(-1) = \rho^{y}_{min}$ and $\tilde{\rho}^{yy}_{i,j}(0) = 0$. A qualified guess is the following function

$$\tilde{\rho}^{yy}_{ij}(\rho^{yy}_{ij}) = a(e^{b\rho^{yy}_{ij}} - 1), \tag{77}$$

clearly $\tilde{\rho}^{yy}_{ij}(0) = 0$ and further

$$\rho^{yy}_{ij,max} = a(e^{b} - 1), \quad \rho^{yy}_{ij,min} = a(e^{-b} - 1), \tag{78}$$

resulting in

$$\frac{\rho^{yy}_{ij,max}}{\rho^{yy}_{ij,min}} = \frac{e^{b} - 1}{e^{-b} - 1} = -e^{b}, \tag{79}$$

where the second equality is found by multiplying by e^b in the denominator and the numerator. Inserting (79) into (78) gives

$$\rho^{yy}_{ij,max} = -a\left(\frac{\rho^{yy}_{ij,max}}{\rho^{yy}_{ij,min}} + 1\right), \quad \Rightarrow \quad a = -\frac{\rho^{yy}_{ij,max}\rho^{yy}_{ij,min}}{\rho^{yy}_{ij,max} + \rho^{yy}_{ij,min}}. \tag{80}$$

Using (68) we can write $\rho^{yy}_{ij,min}$ and $\rho^{yy}_{ij,max}$ as

$$\rho^{yy}_{ij,max} = \frac{e^{\sqrt{\sigma^{zz}_{ii}\sigma^{zz}_{jj}}} - 1}{\sqrt{\sigma^{yy}_{ii}\sigma^{yy}_{jj}}} \hat{Y}_{i,t+1|t}\hat{Y}_{j,t+1|t} \tag{81}$$

$$\rho^{yy}_{ij,min} = \frac{e^{-\sqrt{\sigma^{yy}_{ii}\sigma^{zz}_{jj}}} - 1}{\sqrt{\sigma^{yy}_{ii}\sigma^{yy}_{jj}}} \hat{Y}_{i,t+1|t}\hat{Y}_{j,t+1|t}, \tag{82}$$

with this

$$\rho^{yy}_{ij,max}\rho^{yy}_{ij,min} = \frac{(\hat{Y}_{i,t+1|t}\hat{Y}_{j,t+1|t})^2}{\sigma^{yy}_{ii}\sigma^{yy}_{jj}} \left(2 - e^{\sqrt{\sigma^{zz}_{ii}\sigma^{zz}_{jj}}} - e^{-\sqrt{\sigma^{zz}_{ii}\sigma^{zz}_{jj}}}\right) \tag{83}$$

$$\rho^{yy}_{ij,max} + \rho^{yy}_{ij,min} = \frac{\hat{Y}_{i,t+1|t}\hat{Y}_{j,t+1|t}}{\sqrt{\sigma^{yy}_{ii}\sigma^{yy}_{jj}}} \left(-2 + e^{\sqrt{\sigma^{zz}_{ii}\sigma^{zz}_{jj}}} + e^{-\sqrt{\sigma^{zz}_{ii}\sigma^{zz}_{jj}}}\right). \tag{84}$$

This results in

$$a = \frac{\hat{Y}_{i,t+1|t}\hat{Y}_{j,t+1|t}}{\sqrt{\sigma^{yy}_{ii}\sigma^{yy}_{jj}}} = \frac{1}{\sqrt{(e^{\sigma^{zz}_{ii}} - 1)(e^{\sigma^{zz}_{jj}} - 1)}}. \tag{85}$$

Combining (77), (79) and (85) gives

$$\tilde{\rho}^{yy}_{ij}(\rho^{yy}_{ij}) = \frac{\hat{Y}_{i,t+1|t}\hat{Y}_{j,t+1|t}}{\sqrt{\sigma^{yy}_{ii}\sigma^{yy}_{jj}}} \left(e^{\rho^{yy}_{ij}\sqrt{\sigma^{zz}_{ii}\sigma^{zz}_{jj}}} - 1\right), \tag{86}$$

which is recognized as (75) with $\rho^{zz}_{i,j}$ replaced by $\rho^{yy}_{i,j}$. The transformation in terms of the covariance is

$$\tilde{\sigma}^{yy}_{ij}(\rho^{yy}_{ij}) = \hat{Y}_{i,t+1|t}\hat{Y}_{j,t+1|t} \left(e^{\rho^{yy}_{ij}\sqrt{\sigma^{zz}_{ii}\sigma^{zz}_{jj}}} - 1\right). \tag{87}$$

The transformed covariance matrix for $\boldsymbol{Y}_{t+1}|\mathcal{Y}_t$ becomes

$$\left(\tilde{\boldsymbol{\Sigma}}^{yy}_{t|t+1}\right)_{i,j} = \begin{cases} \sigma^{yy}_{ii} & \text{for} \quad i = j \\ \tilde{\sigma}^{yy}_{ij} & \text{for} \quad i \neq j, \end{cases} \tag{88}$$

and the transformed covariance matrix for $\boldsymbol{Z}_{t+1}|\mathcal{Y}_t$ becomes

$$\tilde{\boldsymbol{\Sigma}}^{zz}_{t+1|t} = \log\left(\tilde{\boldsymbol{\Sigma}}^{yy}_{t+1|t} + \hat{\boldsymbol{Y}}_{t+1|t}\hat{\boldsymbol{Y}}^{T}_{t+1|t}\right) - \log\left(\hat{\boldsymbol{Y}}_{t+1|t}\hat{\boldsymbol{Y}}^{T}_{t+1|t}\right). \tag{89}$$

Denote the innovation of $\boldsymbol{Z}_{t+1}$ by $\tilde{\boldsymbol{Z}}_{t+1|t}$, then

$$\tilde{\boldsymbol{Z}}_{t+1|t} = \log\left(\boldsymbol{Y}_{t+1}\right) - \log\left(\hat{\boldsymbol{Y}}_{t+1|t}\right) + \frac{1}{2}\text{diag}\tilde{\boldsymbol{\Sigma}}^{zz}_{t+1|t}, \tag{90}$$

and the distribution of $\boldsymbol{Z}_{t+1}|\mathcal{Y}_t$ is

$$f(\boldsymbol{Z}_{t+1}|\mathcal{Y}_t) = \frac{1}{\sqrt{(2\pi)^m \det\tilde{\boldsymbol{\Sigma}}^{zz}_{t+1|t}}} \exp\left[-\frac{1}{2}\tilde{\boldsymbol{Z}}^T_{t+1|t}(\tilde{\boldsymbol{\Sigma}}^{zz}_{t+1|t})^{-1}\tilde{\boldsymbol{Z}}_{t+1|t}\right]. \tag{91}$$

Let $\boldsymbol{\theta}$ denote the vector of unknown parameters, then the likelihood function becomes

$$\begin{aligned} L(\boldsymbol{\theta};\mathcal{Y}_N) &= f(\mathcal{Y}_N|\boldsymbol{\theta}) && (92)\\ &= f(\boldsymbol{Z}_N|\mathcal{Y}_{N-1},\boldsymbol{\theta}) f(\boldsymbol{Z}_{N-1}|\mathcal{Y}_{N-2},\boldsymbol{\theta}) \cdots f(\boldsymbol{Z}_1|\boldsymbol{Z}_0,\boldsymbol{\theta}) && (93)\\ &= \prod_{i=1}^{N} \frac{1}{(2\pi)^{\frac{m}{2}}\sqrt{\det\tilde{\boldsymbol{\Sigma}}^{zz}_{i|i-1}}} \exp\left[-\frac{1}{2}\tilde{\boldsymbol{Z}}^T_{i|i-1}(\tilde{\boldsymbol{\Sigma}}^{zz}_{i|i-1})^{-1}\tilde{\boldsymbol{Z}}_{i|i-1}\right], && (94) \end{aligned}$$

and the log likelihood function becomes

$$\log L(\boldsymbol{\theta};\mathcal{Y}_N) = -\frac{1}{2}\sum_{i=1}^{N}\left(\log(\det\tilde{\boldsymbol{\Sigma}}^{zz}_{i|i-1}) + \tilde{\boldsymbol{Z}}^T_{i|i-1}(\tilde{\boldsymbol{\Sigma}}^{zz}_{i|i-1})^{-1}\tilde{\boldsymbol{Z}}_{i|i-1}\right) + \text{const.} \tag{95}$$

Now the ML estimate is found by

$$\hat{\boldsymbol{\theta}} = \arg\left\{\max_{\theta} \log L(\boldsymbol{\theta};\mathcal{Y}_N)\right\}. \tag{96}$$

An example of likelihood estimation is presented in Section 4.1, where the model presented in Section 3.4 is considered.

4.1 Example (Continued)

To illustrate the estimation procedure, we shall now use the described procedure to estimate parameters of the model presented in Section 3.4. In the estimation it is assumed that the structure of the model is known and further that $\boldsymbol{B}$, $\boldsymbol{C}$ and the mean and variance of U_t are known. With this $\boldsymbol{\theta}$ becomes

$$\boldsymbol{\theta} = [a_{1,1}, a_{2,1}, a_{2,2}, \sigma^2_{\xi_1}, \sigma^2_{\xi_2}, \sigma^2_{\lambda_1}, \sigma^2_{\lambda_2}]. \tag{97}$$

The log-likelihood is optimized using the standard optimizer "`optim`" in "R" (see `www.r-project.org`). The method used is "`Nelder-Mead`", which is stable but relatively slow. To ensure that all parameter estimates remain positive, the estimation is done in the log-domain, i.e. we estimate

$$\boldsymbol{\psi} = \log(\boldsymbol{\theta}), \tag{98}$$

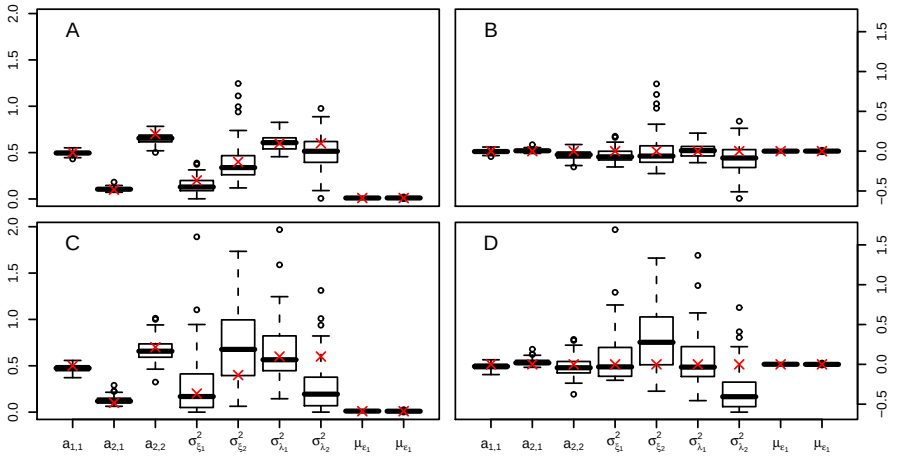

Figure 3: Box plots of estimated parameters ($\hat{\boldsymbol{\theta}}_{LN}$) based on 100 simulated processes of (58)-(60). The boxes show the IQR (Inter Quartile Range), the median is marked with the bold line in the boxes. The whiskers are placed at the most extreme values in the interval [1st quartile- 1.5·IQR, 3rd quartile+1.5·IQR], while more deviating parameter estimates are marked as outliers and plotted with black circles. Left panel shows the actual parameter estimates, while the right panel shows the deviation from the true parameter values. A and B are based on the local log-normal estimation procedure whereas C and D are based on the local Gaussian estimation procedure.

with this we optimize variables from the real line, and these will be approximately Gaussian distributed, and $\boldsymbol{\theta}$ will therefore be log-normal distributed. The initial value of the process is as in Section 3.4.

Since the Gaussian procedure is quite standard and straightforward to implement, we will compare results from the local Gaussian estimation procedure, equations (64)-(65), and the log-normal procedure, equations (95)-(96). The estimation is carried out under the same conditions as for the log-normal estimation.

The merits of the two estimation procedures are each assessed from 100 independent simulations of the model in Section 3.4 by comparing parameter estimates with the true parameter values. The log-normal procedure is generally performing better than the local Gaussian assumption (Figure 3), with smaller variation on parameter estimates and fewer outlier estimates in the simulation experiment. The parameters in $\boldsymbol{A}$ are most precisely estimated with narrow distributions, whereas the estimated multiplicative noise parameters have larger ranges. This is because random variation in the observations can be described by either the

Table 1: Medians of estimated parameters of the process described by equations (58)-(60), based on $M = 100$ independent realizations of the process. Parameters were estimated using the log-normal assumption ($\tilde{\boldsymbol{\theta}}_{LN}$) and the Gaussian assumption ($\tilde{\boldsymbol{\theta}}_{G}$), and multiplicative standard errors were calculated for all parameter estimates. Where the multiplicative standard deviations s_G and s_{LN} are calculated as in eq. (63).

	$\boldsymbol{\theta}$	$\tilde{\boldsymbol{\theta}}_{LN}$	s^2_{LN}	$\tilde{\boldsymbol{\theta}}_{G}$	s^2_{G}
$a_{1,1}$	5.00e−01	4.96e−01	1.09	4.74e−01	1.17
$a_{2,1}$	1.00e−01	1.05e−01	1.40	1.18e−01	1.88
$a_{2,2}$	7.00e−01	6.55e−01	1.18	6.57e−01	1.44
$\sigma^2_{\xi_1}$	2.00e−01	1.30e−01	2.74	1.69e−01	5.91
$\sigma^2_{\xi_2}$	4.00e−01	3.39e−01	2.59	6.94e−01	2.72
$\sigma^2_{\lambda_1}$	6.00e−01	6.08e−01	1.32	5.66e−01	2.33
$\sigma^2_{\lambda_2}$	6.00e−01	5.14e−01	1.97	1.95e−01	4.91
μ_{ϵ_1}	1.00e−02	1.03e−02	1.71	1.12e−02	1.98
μ_{ϵ_2}	1.00e−02	1.10e−02	1.69	9.75e−03	1.98

variances of λ_i or the variances of ξ_i, leading to strong correlations between the estimates of the multiplicative noise terms in the observations and in the system. The general picture is, however, that the estimates are sufficiently accurate, with relatively few estimates far from the true parameter values.

It is seen that the median of the estimates from the 100 simulations is close to the true parameter and that the multiplicative standard deviation is better for the log-normal assumption than the Gaussian assumption for all parameters (Table 1). Even though Table 1 is convincing in terms of which procedure performs the best, it does not allow us to distinguish between the two methods in a statistical sense, i.e. if the parameters from the two approaches are equally good or alternatively if the log-normal estimation approach is significantly better than the Gaussian estimation approach.

To address this question, the estimates obtained from the two different optimization procedures are therefore compared quantitatively in the following. In a simulation study such as the one presented here, it is natural to compare the parameter estimates in a least square sense, i.e. the sum of squared difference between the estimate and true parameter value. In order to do this we collect

Table 2: Loss-functions for parameter estimates based on the two estimation procedures (log-normal and Gaussian), with $M = 200$ independent realizations (100 for each estimation procedure) of the process described by equations (58)-(60), the ratio between the loss-function and an F-test of the hypothesis that the estimations are equally good.

	$S(\hat{\mathbf{\Psi}}_{LN}, M)_j$	$S(\hat{\mathbf{\Psi}}_{G}, M)_j$	$F = \frac{S(\hat{\mathbf{\Psi}}_{G}, M)_j}{S(\hat{\mathbf{\Psi}}_{LN}, M)_j}$	$P\{x < F\}$
$a_{1,1}$	2.08e−03	1.04e−02	5.01	1.000
$a_{2,1}$	2.89e−02	1.36e−01	4.72	1.000
$a_{2,2}$	1.25e−02	3.67e−02	2.93	1.000
$\sigma^2_{\xi_1}$	8.32e−01	1.27e+01	15.24	1.000
$\sigma^2_{\xi_2}$	2.19e−01	5.65e−01	2.58	1.000
$\sigma^2_{\lambda_1}$	1.84e−02	1.83e−01	9.96	1.000
$\sigma^2_{\lambda_2}$	4.23e−01	4.77e+00	11.28	1.000
μ_{ϵ_1}	9.63e−02	1.87e−01	1.94	0.999
μ_{ϵ_2}	9.39e−02	1.96e−01	2.09	1.000

the M independent parameter estimates in the matrix

$$\hat{\mathbf{\Psi}} = \begin{bmatrix} \hat{\psi}_1^T \\ \vdots \\ \hat{\psi}_M^T \end{bmatrix}, \tag{99}$$

and let $\hat{\mathbf{\Psi}}_{LN}$ be the estimates based on the log-normal assumption and $\hat{\mathbf{\Psi}}_{G}$ the estimates based on the Gaussian assumption. The squared distance to the true parameter is calculated as

$$S(\hat{\mathbf{\Psi}}, M)_j = \sum_{i=1}^{M} (\hat{\mathbf{\Psi}}_{i,j} - log(\boldsymbol{\theta}_i))^2. \tag{100}$$

If the parameter estimates are Gaussian random variables with expectation equal to the true parameter, then $S(\hat{\mathbf{\Psi}}, M)_j$ will be χ^2-distributed and $\frac{S(\hat{\mathbf{\Psi}}_{LN}, M)_j/M}{S(\hat{\mathbf{\Psi}}_{N}, M)_j/M}$ will be F-distributed with (M, M) degrees of freedom. Table 2 shows that the loss function for the log-normal estimation procedure ($S(\hat{\mathbf{\Psi}}_{LN}, M)$) is better for all the parameters and that the F-test is significant on a 0.1% level for all the parameters. The conclusion is therefore that the performance of the proposed log-normal assumption is better than the Gaussian assumption.

In the model considered so far the multiplicative standard error for the noise terms ranged between 1.5 and 2, i.e. the density functions are poorly approxi-

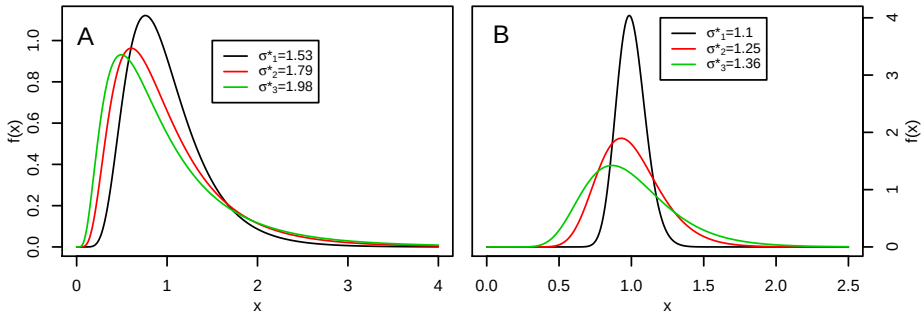

Figure 4: Pdf's for the multiplicative noise terms in the models considered in this section. The multiplicative standard errors (σ_i^*) correspond to variances in the log-domain: A) $\sigma_1^2 = 0.2$, $\sigma_2^2 = 0.4$ and $\sigma_3^2 = 0.6$, and B) $\sigma_1^2 = 0.01$, $\sigma_2^2 = 0.05$ and $\sigma_3^2 = 0.1$.

mated by Gaussian distributions (see Figure 4A). We will compare the performance of the two procedures for a model with smaller multiplicative standard errors and therefore more symmetric density functions. The variances for the multiplicative noise are set to

$$\boldsymbol{\Sigma}_t^{\boldsymbol{\xi}} = \begin{bmatrix} 0.05 & 0 \\ 0 & 0.01 \end{bmatrix}, \quad \boldsymbol{\Sigma}_t^{\boldsymbol{\lambda}} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}, \tag{101}$$

corresponding to multiplicative standard errors between 1.1 and 1.36, while all other parameters are as above. In this case the distribution functions are more symmetric and therefore better approximated by Gaussian density functions (Figure 4B). With this variance parameter setting, the parameters in $\boldsymbol{A}$ and $\boldsymbol{\Sigma}_t^{\boldsymbol{\lambda}}$ as well as μ_{ϵ_1} are more precisely determined by the log-normal procedure on a 95% significance level (Table 3). On the other hand the multiplicative state noise $\boldsymbol{\Sigma}_t^{\boldsymbol{\xi}}$ is better determined by the Gaussian procedure, although not significantly better. A general remark is that the ratio of the loss functions is closer to one for all parameters than was the case in the previous example, meaning that the difference between the procedures is less significant. Furthermore the parameters where the Gaussian procedure performs better than the log-normal procedure are the variance parameters of the noise terms with the lowest multiplicative standard deviation.

Møller et. al. (2009) presents more examples of the filtering and the estimation procedure, the examples have more complex structure than the examples presented in this paper. Further the report prented an examples with negative values of the states but scaled dependent noise. Finally a recipe's on how to deal with missing data is presented. The examples in Møller et. al. (2009)

support the general picture presented in this paper.

Table 3: Loss-functions for parameter estimates based on the two estimation procedures (log-normal and Gaussian), with $M = 200$ independent realizations of the process (100 for each estimation procedure) described by equations (58)-(59) and (101), the ratio between the loss-function and an F-test of the hypothesis that the estimations are equally good.

	$S(\hat{\mathbf{\Psi}}_{LN}, M)_j$	$S(\hat{\mathbf{\Psi}}_{G}, M)_j$	$F = \frac{S(\hat{\mathbf{\Psi}}_{G}, M)_j}{S(\hat{\mathbf{\Psi}}_{LN}, M)}$	$P\{x < F\}$
$a_{1,1}$	4.80e−04	7.48e−04	1.56	0.986
$a_{2,1}$	1.98e−03	3.85e−03	1.95	1.000
$a_{2,2}$	2.08e−04	3.16e−04	1.52	0.981
$\sigma^2_{\xi_1}$	1.42e+00	1.18e+00	0.83	0.179
$\sigma^2_{\xi_2}$	2.17e+00	1.97e+00	0.91	0.314
$\sigma^2_{\lambda_1}$	1.49e−02	2.63e−02	1.77	0.998
$\sigma^2_{\lambda_2}$	1.17e−02	2.60e−02	2.22	1.000
μ_{ϵ_1}	2.19e−02	3.49e−02	1.59	0.989
μ_{ϵ_2}	3.71e−02	3.51e−02	0.94	0.389

5 Discussion and Conclusion

A Kalman filter for random processes with multiplicative noise has been derived using the projection theorem. The filter has proven to work well in the simulation study in that it captures the behavior of the true process well. The filter equations have been developed for general linear processes with multiplicative noise, but the study emphasizes the special case of positive processes and suggests measures to ensure positive reconstructions (and positive predictions).

Furthermore a log-normal maximum likelihood procedure for estimating the parameters of the same family of processes has been derived. This estimation procedure uses the suggested filtering procedure to evaluate the likelihood function. The log-normal maximum likelihood estimation produces seemingly unbiased parameter estimates in the simulation study presented. Maximum likelihood procedures with local Gaussian assumptions are well-known and widely used, but our study has shown that the log-normal likelihood estimation will perform better in cases where the pdf of the multiplicative noise is asymmetric. When the pdf of the multiplicative noise is nearly symmetric, the difference between a Gaussian and a log-normal assumption becomes less pronouced. The log-normal estimation procedure will only work when the process is strictly pos-

itive. For processes that can take negative values, the estimation must be based on other assumptions for the distribution, and in this case the local Gaussian assumption is often the obvious choice.

Acknowledgements

This study is a contribution to the EU-funded project Thresholds (GOCE-003933).

References

Chow B. S. and Birkemeier W. P. (1990) A New Recursive Filter for Systems With Multiplicative Noise, *IEEE Signal Processing Letters*, **14**(7), 469-472

Fenton L. F. (1960) The Sum of Log-Normal Probability Distributions in Scatter Transmission Systems, *IEEE Transaction and Communication*, **8**(1), 57-67

Jazwinski A. H. (1970) Stochastic Processes and Filtering Theory, *Academic Press*, New York.

Jimenez J. C. and Ozaki T. (2002) Linear Estimation of Continuous-Discrete Linear State Space Models with Multiplicative Noise, *System & Control Letters*, **47**, 91-101

Kalman R. E. (1960) A New Approach to Linear Filtering and Prediction Problems, *Transactions of the ASME-Journal of Basic Engineering, Series D*, **82**, 35-45

Kotz S., Balakrishnan N. and Johnson L. N. (2000) Continuous Multivariate Distributions, Volume 1: Models and Applications, *Wiley Series in Probability and Statistics*

Kristensen N.R., Madsen H. and Jørgensen S.B. (2004) Parameter estimation in stochastic grey-box models, *Automatica*, **40**, 225-237

Limpert E., Stahel W. A. and Abbt Markus (2001) Log-normal Distributions across the Sciences: Keys and Clues, *BioScience*, **51**(5),341-352

Guttal V., Jayaprakash C. Impact of noise on bistable ecological systems. *Ecological Moddeling*, **201**(3), 420-428

Madsen, H. (2008) Time Series Analysis, *Chapman & Hall*

Møller J. K., Carstensen J., Madsen H. and Andersen T. (2008) Dynamic two State Stochastic Models for Ecological Regime Shifts, *Environmetrics*, **20**, 912-927

Møller J. K., Carstensen J., Madsen H. (2009) Parameter Estimation in State Space Models with Multiplicative Noise - Examples, *IMM-Technical Report-2009-09* Informatics and Mathematical Modelling, Technical University of Denmark `URL: http://orbit.dtu.dk/All.external?recid=240477`

Nielsen J.N., Madsen H., Young P.C. (2000). *Parameter Estimation in Stochastic Differential Equations; An Overview*, Annual Reviews in Control, **24**, pp. 83-94.

Overgaard R. V., Jonsson N., Tornøe C. and Madsen H. (2005) Non-Linear Mixed-Effects Models with Stochastic Differential Equations: Implementation of an Estimation Algorithm, *Journal of Pharmacokinetics and Pharmacodynamics*, **32**(1)

Vio R., Kristensen N.R., Madsen H. and Wamsteker W. (2005). *Time Series Analysis in Astronomy; Limits and Potentialities*, Astronomy and Astrophysics, Vol. 435, pp. 773-780

Yang F., Wang Z. and Hung Y.S. (2002) Robust Kalman Filtering for Discrete Time-Varying Uncertain Systems With Multiplicative Noises, *IEEE Transaction on Automatic Control*, **47**(7),1179-1183

Zhang L. and Zhang X.-D. (2007) An Optimal Filtering Algorithm for Systems With Multiplicative/Additive Noises, *IEEE Signal Processing Letters*, **14**(7),469-472

Øksendal B. (2003) Stochastic Differential Equations - An Introduction with Applications, Sixth Edition, *Springer*

Paper C

Parameter Estimation in a Simple Stochastic Differential Equation for Phytoplankton Modelling

---

Authors:
J.K. Møller, H. Madsen, and J. Carstensen.

Submitted to:
Ecological Modelling, **222**, 1793-1799, (2011).

Parameter Estimation in a Simple Stochastic Differential Equation for Phytoplankton Modelling

Jan Kloppenborg Møller[1,2], Henrik Madsen[1],
Jacob Carstensen[2]

Abstact

The use of stochastic differential equations (SDEs) for simulation of aquatic ecosystems has attracted increasing attention in recent years. The SDE setting also provides the opportunity for statistical estimation of ecosystem parameters. We present an estimation procedure, based on Kalman filtering and likelihood estimation, which has proven useful in other fields of application. The estimation procedure is presented and the development from ordinary differential equations (ODEs) to SDEs is discussed with emphasis on autocorrelated residuals, commonly encountered with ODEs. The estimation procedure is applied to a simple nitrogen-phytoplankton model, with data from a Danish estuary (1988-2006). The resulting SDE is simple enough to have a well-known stationary distribution and this distribution is presented and compared with observed phytoplankton data.

Key Words: Phytoplankton modelling, Stochastic differential equations, parameter estimation, Extended Kalman Filter, Maximum likelihood estimation.

1 Introduction

Ecosystems, and marine ecosystems in particular, are complex mosaics of interconnected processes, with many known and unknown drivers affecting the systems. Marine ecosystems have traditionally been modelled by means of Ordinary Differential Equations (ODEs) that have gradually evolved from simple descriptions of the nutrient-phytoplankton interaction (NP models) to include

[1]DTU Informatics, Richard Pedersens Plads, Technical University of Denmark Building 321, DK-2800 Lyngby, Denmark.
[2]National Environmental Research Institute, Fredriksborgvej 399, DK-4000 Roskilde, Denmark.

an increasing number of components, e.g. zooplankton (NPZ models), detritus (NPZD models), benthic fauna and vegetation, as well as fish. Moreover, nutrients and organisms have also gradually been partitioned into different constituents and species groups, often in response to inadequate description of observed dynamics. The consequence of employing such a reductionistic approach is increasing model complexity with an escalating number of unknown parameters that are calibrated using values obtained from the literature or tuned to mimic observations, which are often aggregates of several states, according to the modeller's subjective assessment. For example, Fasham et al. (1990) considered a relatively simple NPZD model with 7 states (3 different nitrogen pools, phytoplankton, zooplankton, detritus and bacteria) and 25 parameters, whereas Bartell et al. (1999) introduced a flexible, though complex, modelling framework and applied it to Canadian lakes using 44 states for the biological components. Ecological models seem to have grown in size and complexity as computational constraints have been alleviated and understanding of subprocesses have grown over time. Matear (1995) employ and more objective stochastic optimisation (simulated annealing), but the underlying system is still deterministic.

Scientists have come to realize that even the most complex ecosystem model will not be able to capture all mechanisms and drivers of the real ecosystem, Dowd (2006, 2007) presents NPD models with differential random forcing, the formulation of the noise does however not allow a stochastic differential equation formulation (Øksendal (2003)).

Drivers that are unobservable or not accounted for in a model will lead to systematic deviations from the model in the form of autocorrelated residuals between observations and short term predictions. This kind of autocorrelation is actually also evident in the results presented by Fasham et al. (1990) (Figures 4 and 5 in the reference). In fact, these residuals can be modelled as stochastic perturbations working within the model. In this case, ODE models with stochastic input of internal randomness are referred to as Stochastic Differential Equations (SDEs) (e.g. Øksendal (2003)). This internal stochastic perturbation will also, to some extent, be able to indirectly capture drivers of the system not implicitly contained in the model formulation.

SDEs are an emerging field and the use of SDEs in mathematical finance and option pricing (e.g. Black and Scholes models) is the standard example in many text books, (e.g. Øksendal (2003)). Early applications of SDEs in other fields can be found in Madsen (1987) (climatology), Madsen & Holst (1995) (engineering), Jacobsen & Madsen (1996) (oxygen level in a stream). More recently SDEs have been applied to pharmaceutical problems e.g. Tornøe et al. (2004). Modelling of motion patterns for larger animals has also been the subject of SDE-modelling in recent years e.g. Brillinger et al. (2002) and Pedersen et al. (2008).

The use of stochastic differential equations (SDEs) to introduce stochastic forcing in NP-like ecosystem models has attracted increasing attention over recent years. Carpenter & Brock (2006) and Guttal & Jayaprakash (2008) are examples, where SDEs are used in analysis of non-linear stochastic systems with emphasis on regime shifts, i.e. these studies analyse known deterministic regime shift models with stochastic forcing. These studies are, however, pure simulation studies of how random behaviour affects the dynamics of regime shift models. Stollenwerk et al. (2001) present estimation, of phytoplankton in an SDE-based model, based on stationary distribution and consequently on data from the growth season only. The present study use data and SDEs for NP modelling and parameter estimation based on 19 years of data including the winter period. The aim is to provide a simple example to illustrate the usefulness of SDEs (as presented in Øksendal (2003) and Klebaner (2005)) in ecological modelling, as SDEs have proven useful in other fields of application.

The paper is organised in the following way: Section 2 gives a short introduction to SDEs, focusing on the development from ODEs to SDEs, Section 3 introduces the statistical method used for parameter and state estimation in SDEs, Section 4 presents a real-life example of a simple NP-SDE model with data from an estuary in the northern Denmark. The SDE model includes phytoplankton as the state with water-column nitrogen and global radiation (total (i.e. direct and diffuse) incoming solar radiation) as drivers.

2 Stochastic Differential Equations and Ordinary Differential Equations

As a general rule it is only possible to observe continuous time processes in discrete time. Let $\boldsymbol{x}_t \in \mathcal{X} \subset \mathbb{R}^n$ be the continuous time state variable which is observed through an observation equation in discrete time, and let $\boldsymbol{y}_k \in \mathcal{Y} \subset \mathbb{R}^l$ denote the observation at time t_k ($k \in \{0, ..., N\}$), let the observation equation be given by

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k, \boldsymbol{u}_k, t_k, \boldsymbol{\theta}, \boldsymbol{e}_k), \tag{1}$$

where $\boldsymbol{x}_k$ and $\boldsymbol{u}_k \in \mathcal{U} \subset \mathbb{R}^r$ is the state variable and the inputs (forcing or control variables) at time $t = t_k$, $\boldsymbol{e}_k \in \mathbb{R}^l$ is the random observation error, $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ is a set of parameters to be estimated and $\boldsymbol{h}(\cdot) \in \mathbb{R}^l$ is the function that links the states to the observations. Simple forms of $\boldsymbol{h}(\cdot)$ include the identity link ($\boldsymbol{h}(\cdot) = \boldsymbol{x}_{t_k} + \boldsymbol{e}_k$) and the loglink ($\boldsymbol{h}(\cdot) = \log(\boldsymbol{x}_{t_k}) + \boldsymbol{e}_k$ (if $\boldsymbol{x}_t > \boldsymbol{0}$)), with $\boldsymbol{e}_k \sim N(0, \boldsymbol{S}_k)$ and $N(\cdot)$ is the normal distribution.

2.1 Ordinary Differential Equation representation

In the ordinary differential equation setting the evolution in time of the state variable is given by the deterministic system equation

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})dt, \tag{2}$$

where $t \in \mathbb{R}$ is time, (the structure of) $\boldsymbol{f}(\cdot) \in \mathbb{R}^n$ is deduced from physical (or biological) knowledge of the system, and $\boldsymbol{u}_t$ and $\boldsymbol{\theta}$ are similar to the input and parameters presented in the observation equation (1).

If $\boldsymbol{e}_k$ takes a simple form (i.e. additive and Gaussian) and $\boldsymbol{x}_t$ follows the deterministic formulation (2), then the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is equivalent to minimising the weighted sum of squared observation errors ($\boldsymbol{e}_k^T \boldsymbol{S}_k^{-1} \boldsymbol{e}_k$), where $\boldsymbol{S}_k$ is the observation covariance matrix for the k'th observation.

2.2 Stochastic Differential Equation representation

Real-life systems are subject to random perturbation, such as random variation of the input (specified by $\boldsymbol{u}_t$) or non-specified random forcing, e.g. processes not specified in the model description, working within the system. Such perturbations create autocorrelated noise in the observations ($\boldsymbol{y}_k$), which cannot be captured by Eqs. (1) and (2), since observation noise is present only. Further, when the parameters, $\boldsymbol{\theta}$ in Eq. (2), have been estimated then all future values of the system are known independent of the forecast horizon, which is somewhat counterintuitive.

SDEs can be formulated by introducing a noise term, perturbing the differential of $\boldsymbol{x}_t$ (Øksendal (2003))

$$\frac{d\boldsymbol{x}_t}{dt} = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta}) + \boldsymbol{\sigma}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})\boldsymbol{w}_t, \tag{3}$$

where $\boldsymbol{w}_t \in \mathbb{R}^m$ is an m-dimensional standard Wiener process and $\boldsymbol{\sigma}(\cdot) \in \mathbb{R}^{n \times m}$ is a matrix function (Øksendal (2003)). Multiplying with dt gives the standard SDE formulation

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})d\boldsymbol{w}_t, \tag{4}$$

$\boldsymbol{\sigma}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})$ is referred to as the diffusion term, and $\boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})$ is referred to as the drift term. The solution to (4) is a stochastic process with transition probabilities given by the Fokker-Planck equation (e.g. Klebaner (2005)). Furthermore, one path of the solution is an autocorrelated stochastic process, which can be realized by considering Eq. (3) where the increments of $\boldsymbol{x}_t$ are subject to random perturbations.

2.3 A simulation example

A simple example, that resembles some features of the case study presented in Section 4, can illustrate the points discussed above. Consider the SDE

$$dx_t = \left[\sin\left(\frac{2\pi}{12}t\right) + 1 - ax_t\right] dt + \sigma x_t dw_t, \tag{5}$$

with $a = 0.5$ and $\sigma = 0.2$, x_t can be considered as an ecosystem component (e.g. phytoplankton) with a periodic growth process $\left(\sin\left(\frac{2\pi}{12}t\right) + 1\right)$ which is independent of the state, a death-rate (a), and a diffusion term that is proportional to the state of the system. The solution to $dx_t = -ax_t dt + \sigma x_t dw_t$, is a log-normal distributed random variable if the initial state (x_0) is larger than zero (e.g. Øksendal (2003) (Example 5.1.1)). Therefore adding a positive forcing, will still guarantee that $x_t > 0 \; \forall \; t$, if $x_0 > 0$. Assume that the observation equation is given by

$$\log(y_k) = \log(x_{t_k}) + e_k, \tag{6}$$

with $e_k \sim N(0, 0.1^2)$, this implies that the standard deviation of the observation noise is proportional to the state of the system. As discussed above, the ODE solution (Figure 1) is a deterministic function, with autocorrelated residuals. The SDE solution is a stochastic process represented by the expectation (Figure 1) and covariance (not shown) of the state given all observations. Clearly this expectation captures the autocorrelation of the underlying process quite well (Figure 1).

The example demonstrate that the SDE solution captures the dynamics of the underlying process better than the ODE solution (the residual sum of square is 0.18 and 1.60 for the SDE solution and the ODE solution respectively). When SDEs are used for long-term forecasts autocorrelation between the residuals will be observed, since x_t is an autocorrelated process, e.g. Madsen (2008) for this result in linear time series analysis. However the distributional properties of long-term forecasts will be captured better with SDEs than with ODEs.

3 Parameter Estimation in SDEs

Estimation of parameters in SDEs is a difficult task because evaluation of the likelihood of observation requires knowledge about the transition densities between discrete time observations. Transitions densities are generally unknown except for very simple SDEs and approximate methods has to applied. To enable general estimation of SDEs simulation based methods has to be applied (e.g.

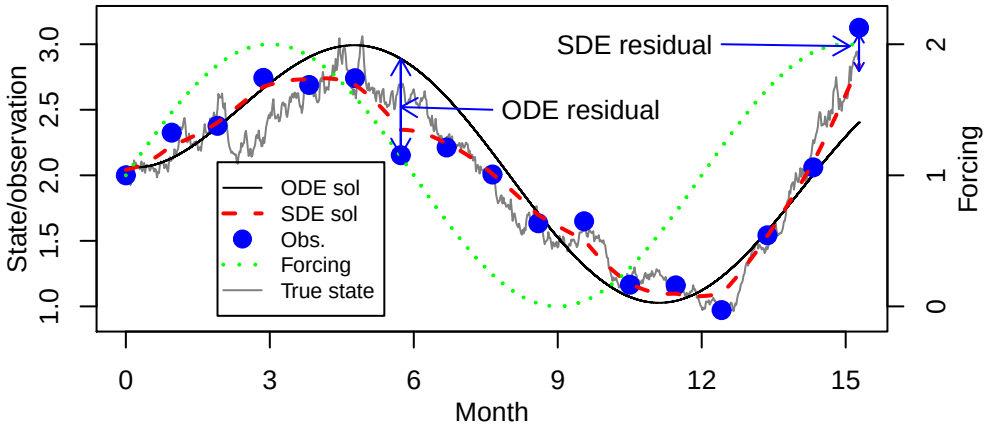

Figure 1: Simulation results from the system described in Eqs. (5) and (6).

Nicolau (2002)), including sampling techniques (e.g Pastorello & Rossi (2010)) and particle filters (e.g. Givon et al. (2009)). While simulation based method has the advantage of dealing effectively with general differential noise terms like Possion noise (Givon et al. (2009)), the draw back is the computational effort needed in the simulation part of the algorithms.

Closed form likelihood expansions are also available (e.g. Aït-Sahalia (2008)), while tractable from a computational point of view, these are complicated to apply involving Hermite series expansion of the likelihood. Further the assumption is that all states are observed, which might not be the case in real life examples. In the present study we will base the analysis on the Extended Kalman filter (EKF), while well-known (Kristensen et al. (2004)) the method has, to our knowledge, not been applied to Marine ecosystems previously. A key advantage of the methodology is that the method is implemented in the easily accessible open source software CTSM[1] (Kristensen & Madsen (2003), Kristensen et al. (2004)).

[1]The software is available at `www2.imm.dtu.dk/~ctsm`

3.1 Likelihood estimation by EKF

Consider the continuous-discrete time stochastic state-space model formulation in Eqs. (1) and (4)

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})d\boldsymbol{w}_t \tag{7}$$

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_{t_k}, \boldsymbol{u}_{t_k}, t_k, \boldsymbol{\theta}, \boldsymbol{e}_k), \tag{8}$$

where parameters and variables are as described in Section 2.

For the continuous-discrete time stochastic state-space model (7) - (8), the problem that needs to be solved is: Find the set of parameters ($\hat{\boldsymbol{\theta}}$) such that some objective function is maximised given a set of observations $\mathcal{Y}_N = \{\boldsymbol{y}_0, ..., \boldsymbol{y}_N\}$. A natural choice of such an objective function is the joint probability density of the observations, considered as a function of the unknown parameters ($\boldsymbol{\theta}$) (the likelihood function), i.e.

$$\begin{aligned} L(\boldsymbol{\theta}; \mathcal{Y}_N) &= p(\mathcal{Y}_N|\boldsymbol{\theta}) \\ &= \left(\prod_{k=1}^{N} p(\boldsymbol{y}_k|\mathcal{Y}_{k-1}, \boldsymbol{\theta}) \right) p(\boldsymbol{y}_0|\boldsymbol{\theta}), \end{aligned} \tag{9}$$

where Bayes rule has been applied recursively to form the product of conditional densities (e.g. Madsen (2008)). In principle the solution of this problem would be an application of the Fokker-Planck equation for predictions and Bayes rule for updating given a new observation. Such a strategy is, however, infeasible, except for systems with very simple structures of the system equation (7), because it involves the solution of a very complex partial differential equation.

The estimation procedure, which will be introduced in the following, relies on an implementation of EKF techniques (Jazwinski (1970)). This implementation requires the system and observation equation to have the form

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{u}_t, t, \boldsymbol{\theta})d\boldsymbol{w}_t \tag{10}$$

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_{t_k}, \boldsymbol{u}_{t_k}, t_k, \boldsymbol{\theta}) + \boldsymbol{e}_k, \tag{11}$$

where 1) the diffusion matrix is quadratic, i.e. $\boldsymbol{\sigma}(\cdot) \in \mathbb{R}^{n \times n}$ and $\boldsymbol{w}_t \in \mathbb{R}^n$, 2) the diffusion term is not allowed to depend on the state, and 3) the observation noise is additive Gaussian white noise ($\boldsymbol{e}_k \in N(\boldsymbol{0}, \boldsymbol{S}_k(\boldsymbol{\theta}, \boldsymbol{u}_k))$). In a weak solution sense (equality in distribution) (see Øksendal (2003) for a discussion of weak and strong solutions), 1) is not a restriction since $\boldsymbol{\sigma}(\cdot)$ can only be identified up to the "square root" of $\boldsymbol{\sigma}\boldsymbol{\sigma}^T(\cdot)$, 2) is clearly a restriction in the multivariate case, but to some extent this can be dealt with by transformations, and a class of diffusion processes can be dealt with by transformations (e.g. all processes

with diffusion given by $\text{diag}(\boldsymbol{x}_t)\boldsymbol{\sigma}(t,\boldsymbol{\theta})$ (Luschgy & Pagés (2006))). Finally, the restriction 3) should be dealt with by transformation of the observation if possible.

Since the system (10)-(11) is driven by Wiener noise, and the observation noise is additive Gaussian, a reasonable local approximation of the conditional densities in (9) is the Gaussian distribution, which is completely characterised by its mean and covariance.

The one-step prediction, covariance and the innovation are defined as

$$\hat{\boldsymbol{y}}_{k|k-1} = E\{\boldsymbol{y}_k|\mathcal{Y}_{k-1},\boldsymbol{\theta}\} \tag{12}$$

$$\boldsymbol{R}_{k|k-1} = V\{\boldsymbol{y}_k|\mathcal{Y}_{k-1},\boldsymbol{\theta}\} \tag{13}$$

$$\boldsymbol{\epsilon}_k = \boldsymbol{y}_k - \hat{\boldsymbol{y}}_{k|k-1}, \tag{14}$$

where $E\{\cdot\}$ is the expectation and $V\{\cdot\}$ is the variance. Using this notation the likelihood can be written as

$$L(\boldsymbol{\theta};\mathcal{Y}_N) = \left(\prod_{k=1}^{N} \frac{\exp(\frac{1}{2}\boldsymbol{\epsilon}_k^T \boldsymbol{R}_{k|k-1}^{-1}\boldsymbol{\epsilon}_k)}{\sqrt{\det\left(\boldsymbol{R}_{k|k-1}\right)(2\pi)^l}}\right) p(\boldsymbol{y}_0|\boldsymbol{\theta}), \tag{15}$$

where l is the dimension of the sample space (see Eq. (1)) and $(\cdot)^T$ is the vector transpose. The actual optimisation is done in the log-domain and the (approximate) maximum likelihood estimate of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta}\{\log(L(\boldsymbol{\theta};\mathcal{Y}_N))\}. \tag{16}$$

The Kalman gain is essential for the state updating procedure. The Kalman gain governs how much the one-step prediction ($\hat{\boldsymbol{x}}_{k|k-1}$) should be adjusted to form the reconstruction ($\hat{\boldsymbol{x}}_{k|k}$) of the state based on the observation, and is given by

$$\boldsymbol{K}_k = \boldsymbol{P}_{k|k-1}\boldsymbol{C}^T\boldsymbol{R}_{k|k-1}^{-1}, \tag{17}$$

where $\boldsymbol{C}$ is the first order Taylor expansion (the Jacobian) of $\boldsymbol{h}$ and $\boldsymbol{P}_{k|k-1}$ is the covariance of the one-step prediction. Note that the Kalman gain proportional to the information ($\boldsymbol{R}_{k|k-1}^{-1}$) provided by the k'th observation. The state reconstruction is given by

$$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k\boldsymbol{\epsilon}_k \tag{18}$$

$$\boldsymbol{P}_{k|k} = \boldsymbol{P}_{k|k-1} - \boldsymbol{K}_k\boldsymbol{R}_{k|k-1}^{-1}\boldsymbol{K}_k^T, \tag{19}$$

i.e. a combination of the predicted state and the information obtained by the k'th observation ($\boldsymbol{y}_k$). The state predictions are governed by a set of ordinary differential equations, (Kristensen et al. (2004), Kristensen & Madsen (2003)).

In addition to the state reconstruction and parameter estimates discussed above, the optimisation of the likelihood function provides an estimate of the parameter covariance given by the negative inverse Hessian of the log-likelihood evaluated at the optimal parameter values. As the estimation is based on the maximum likelihood, the procedure allows for likelihood ratio tests of nested models and t-tests for all estimated parameters.

In the simulation example of Section 2.3 the estimated death-rate was 0.508 ($\pm$0.14) and 0.536 ($\pm$0.04) for the SDE model and the ODE model, respectively. Thus, the ODE solution deviates substantially from the true value (0.5) and the 95% confidence interval does not even contain the true value, indicating that the estimate of the dynamics of the process is also better when taking the correlation structure into account.

4 Skive Fjord case study

This section considers a real-life application of the presented estimation techniques. The example is a very simple model for phytoplankton dynamics in an estuary located in the northern Denmark. The model aims at describing phytoplankton nitrogen ($X_{p,t}$) as a function of total nitrogen in the water column ($U_{w,t}$) and incoming global radiation ($U_{gr,t}$). The period with overlapping time series of input data is September 18th 1987 through December 18th 2006, which will be the modelling period.

4.1 Data

Skive Fjord has been extensively monitored during the Danish National Aquatic Monitoring and Assessment Program (DNAMAP), where various ecosystem component and water-chemistry variables have been recorded since the 1980s.

The data set includes chlorophyll in μg chla/l, which is converted to nitrogen units using the standard chlorophyll to carbon ratio of 1:50 (weight), the Redfield ratio (C:N=106:16 (molar)), and assuming that the monitoring station is representative of the entire estuary (the observation is denoted $Y_{p,t}$). Because nitrogen in the water column acts as an input to the system, missing obser-

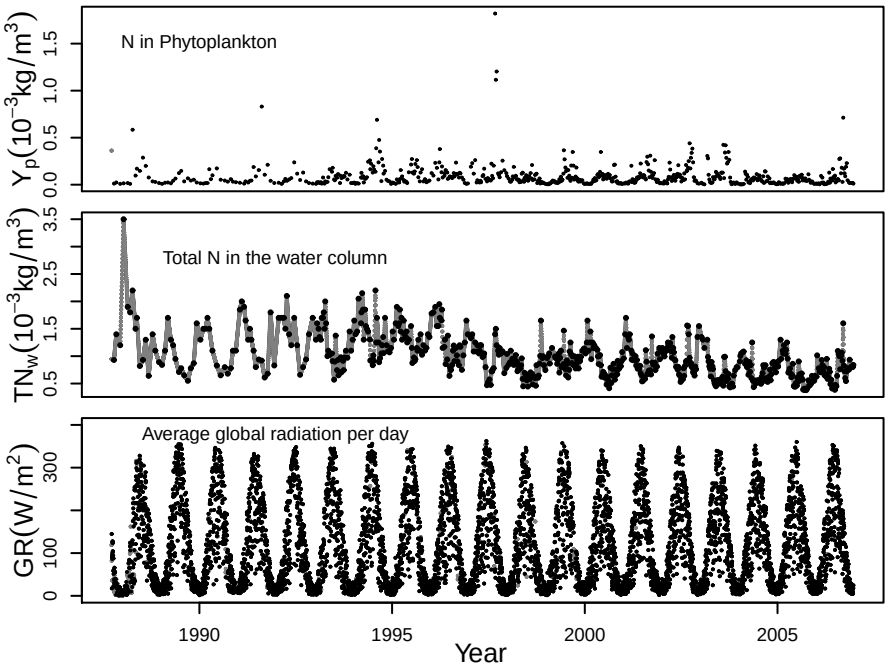

Figure 2: Observations of total nitrogen in phytoplankton, total nitrogen in the water-column and global radiation. Black dots indicate observations, while gray dots/lines indicate observations that are filled in by interpolation.

vations are not allowed and are filled in by linear interpolation between data points.

Global radiation data (provided by the Danish Meteorological Institute) are available from two sites around Skive Fjord and reported on an hourly basis. The global radiation data contains both missing observations and what we will refer to as "false zeros". A false zero is when global radiation equal to zero is reported during daytime. To identify such points, a general yet simple periodic function for global radiation is fitted to the non-missing data

$$f_{gr}(t;\boldsymbol{\theta}) = \left(a_0 + a_1 \sin\left(\frac{2\pi}{P_{year}}t + \phi_1\right) + a_2 \sin\left(\frac{2\pi}{P_{day}}t + \phi_2\right)\right)_+, \qquad (20)$$

where $(x)_+ = \max(x, 0)$, $P_{year} = 24 \times 365.25$ is the average number of hours in one year and $P_{day} = 24$ is the number of hours in one day. If the observation at time t_{i_0} is zero and $f_{gr}(t_i) > 0$ for $i \in \{i_0 - 1, i_0, i_0 + 1\}$ then the observation is considered a false zero and marked as missing. The number of observations

removed in this way is 144 out of a total of about 270×10^3 observations.

The hourly global radiation is found as a simple average (over stations) of the non-missing observations at each time point. As global radiation acts as an input, missing observations are not allowed, (the number of missing observations is about 1% of the total number of observations). If the sequence of missing data is shorter than three, or the same hour of the day before and after is available, then the missing observations are filled in by linear interpolation. Missing data not filled in by linear interpolation is filled in by equating with $f_{gr}(t)$, and $U_{gr,t}$ is created by average daily global radiation.

The seasonal variation in both input variables and phytoplankton nitrogen is evident (Figure 2), but strong fluctuations overlaying this signal are apparent.

4.2 A simple SDE-model

The simple model set up is a constant mortality rate and a growth process which is a function of available nitrogen and global radiation

$$\frac{dX_{p,t}}{dt} = b(U_{w,t}, U_{gr,t}) - aX_{p,t} + noise, \tag{21}$$

where the growth process $b(U_{w,t}, U_{gr,t})$ and the mortality rate (a) are both strictly positive. The growth process will be assumed proportional to the available nitrogen and the inflow of solar energy, i.e.

$$b(U_{w,t}, U_{gr,t}) = b_0 U_{w,t} U_{gr,t}, \tag{22}$$

where $b_0 > 0$ is a constant. This formulation governs a stochastic process and a natural requirement for the process is that $P(X_{p,t} < 0) = 0 \quad \forall t$. A formulation that meets this constraint is

$$dX_{p,t} = (b_0 U_{w,t} U_{gr,t} - aX_{p,t})dt + \sigma_x X_{p,t} dw_t, \tag{23}$$

where w_t is the standard Wiener process. In this formulation the diffusion for the process is proportional to the level of the process, i.e. the higher the abundance of phytoplankton the higher the variance (in absolute terms). This choice of noise process is the simplest in terms of estimation, because the state space of the Lamperti transformed process is the entire real axis, but it is also in good agreement with the generality of the log-normal distribution (Limpert et. al. (2001)) and coincide with the choice in Dowd (2006). Clearly this model is extremely simple and is therefore a coarse simplification of the real system, however it will suffice for illustrating the method.

In general closed form solutions of SDEs are hard or impossible to find, and (23) is no exception (e.g. Iacus (2008); Pastorello & Rossi (2010)). If $b(U_{w,t}, U_{gr,t})$ is constant, then Eq. (23) is a special case of the Pearson diffusion, which is defined as (Iacus (2008))

$$dX_t = -\theta(X_t - \mu)dt + \sqrt{2\theta(\sigma_1 X_t^2 + \sigma_2 X_t + \sigma_3)}dw_t, \tag{24}$$

which does not have a closed form solution for the transient. For $\sigma_2 = \sigma_3 = 0$, $\theta > 0$, $\mu > 0$, and $\sigma_1 > 0$, the stationary distribution for this process is, however, known to be an inverse Gamma distribution with shape parameter $1 + \sigma_1^{-1}$ and scale parameter μ/σ_1 (see Iacus (2008), observe the misprint on p. 54 in the reference though). With $b(\cdot)$ a function of t (the inputs vary in time), the process will always be in the transient, nonetheless the process should be close to the stationary distribution, if the time constant for the system $(1/a)$ is fast compared to the variation in $b(\cdot)$. Suppressing the time index and comparing (23) and (24) gives

$$\theta = a, \quad \mu = \frac{b_0 U_w U_{gr}}{a}, \quad \sigma_1 = \frac{\sigma_x^2}{2a}, \tag{25}$$

which implies that for each fixed t the stationary distribution for $X_{p,t}$ follows an inverse gamma distribution with shape parameter $1 + \frac{a}{\sigma_x^2}$ and scale parameter $\frac{2b_0 U_{w,t} U_{gr,t}}{\sigma_x^2}$.

As discussed in Section 3 the diffusion term is required to be independent of the state of the system, which is not the case for the presented system. Fortunately it is always possible to transform a one dimensional system with only one continuously differentiable diffusion term into a system with constant diffusion, by applying the Lamperti transform (e.g. Baadsgaard et al. (1997); Iacus (2008))

$$\psi(X_t) = \int \frac{1}{\sigma(\xi)} d\xi \Big|_{\xi = X_t}, \tag{26}$$

let ψ_t, ψ_x and ψ_{xx} denote $\frac{\partial \psi}{\partial t}$, $\frac{\partial \psi}{\partial x}$ and $\frac{\partial^2 \psi}{\partial^2 x}$ respectively. By choosing $Z_t = \psi(X_{p,t}) = \int \frac{1}{\sigma_x \xi} d\xi \big|_{\xi = X_t} = \frac{1}{\sigma_x} \log(X_{p,t})$, and applying Ito's lemma (e.g. Øksendal (2003)) we get

$$\begin{aligned}
dZ_{p,t} =& \psi_t(X_{p,t})dt + \psi_x(X_{p,t})dX_{p,t} + \psi_{xx}(X_{p,t})(dX_{p,t})^2 && (27)\\
=& \frac{(b_0 U_{w,t} U_{gr,t} - a X_{p,t})dt + \sigma_x X_{p,t} dw_t}{\sigma_x X_{p,t}} - \frac{1}{2}\frac{(\sigma_x X_{p,t})^2}{\sigma_x (X_{p,t})^2} dt && (28)\\
=& \left(\frac{b_0 U_{w,t} U_{gr,t}}{\sigma_x X_{p,t}} - \frac{a}{\sigma_x} - \frac{1}{2}\sigma_x \right) dt + dw_t && (29)\\
=& \left(\frac{b_0}{\sigma_x} e^{-\sigma_x Z_{p,t}} U_{w,t} U_{gr,t} - \frac{a}{\sigma_x} - \frac{1}{2}\sigma_x \right) dt + dw_t, && (30)
\end{aligned}$$

which is now a non-linear SDE with unit diffusion. In addition to the system equation described above, there has to be a description of the observation equation. Under the assumption that observations are log-normal distributed around the true state, the observation equation is

$$\log\left(Y_{p,k}\right) = \sigma_x Z_{p,t_k} + e_k, \tag{31}$$

where $Y_{p,k}$ is the observed nitrogen content in phytoplankton and $e_k \sim N(0, \sigma_y^2)$.

4.3 Results

The parameters of the model Eqs. (30)-(31) are now estimated using the estimation procedure presented in Section 3. All parameters score well in t-tests (Table 1). The time constant $(1/a)$ of the deterministic skeleton (remove the noise term) is about 67 days. As discussed above, the stationary distribution is known when the forcing $(b_0 U_{w,t} U_{gr,t})$ is constant, which is clearly (Figure 2) not the case for the system analysed here. It is, however, reasonable to assume that the state (phytoplankton nitrogen) is close to the stationary distribution in some sense. To explore this we define a moving average growth process

$$\tilde{b}_t = \frac{b_0}{14} \sum_{i=t-13}^{t} U_{w,i} U_{gr,i}. \tag{32}$$

This moving average growth process is now used (in Eq. (25)) to calculate confidence intervals around the mode of the stationary distribution of each t (Figures 3 and 4).

Table 1: Estimation results

	$[\theta_{min}, \theta_{max}]$	$\hat{\theta}$	Std. dev.	t-score	$P(x > \lvert t \rvert)$
$Z_{p,0}$	[-20 , 20]	-9.4e+00	3.1e+00	-3.0	0.003
b_0	[0 , 1]	1.9e-03	2.2e-04	8.4	0.000
a	[0 , 1]	1.7e-02	3.4e-03	5.0	0.000
σ_x	[0 , 1]	1.6e-01	1.3e-02	12.0	0.000
σ_y	[0 , 1]	1.9e-01	2.6e-02	7.6	0.000

In addition to parameter estimates, CTSM allows us to calculate the smoothed state (the most probable state given all observations), i.e.

$$\hat{\boldsymbol{x}}_{t|T} = E[\boldsymbol{x}_t | \mathcal{Y}_N]. \tag{33}$$

The smoothed state is in general quite close to the mode of the stationary distribution (Figures 3 and 4). While Figure 3 gives the distribution along

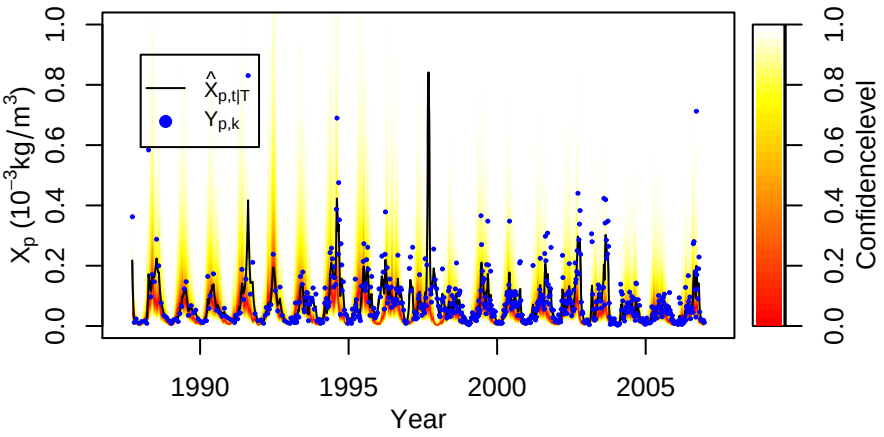

Figure 3: Time series of stationary distributions, smoothed state (black line) and observation (blue dots). The colour key refers to confidence intervals around the mode of the stationary distributions.

the entire time span of the observations, Figure 4 provides a zoom on the years 1999-2001. Again it is evident that the stationary distribution fits the smoothed state quite well, but also that the extremes are not captured very well by the stationary distribution.

Very large observations influence the smoothed state of the system strongly (Figure 3), implying that observations in the tail of the stationary distribution draw the smoothed state away from the centre of the stationary distribution. This indicates that the model does not capture the extremes in the dynamics very well, which is not surprising since the model does not include any mechanism to capture extreme events. It might possible to capture such behaviour by more effects (e.g. $b(t) = b(U_{w,t}, U_{gr,t}, X_{p,t})$) in the system equations.

5 Discussion

We have demonstrated that the presented approach based on embedded stochastic differential equations provides a strong tool for modelling. In particular the procedure accounts for the autocorrelated residuals normally seen when ODEs are used for modelling. Furthermore, as the model is formulated in continuous time, the states can be updated and parameters estimated from data that is not sampled at equidistant points in time, which is often the case for ecosystem

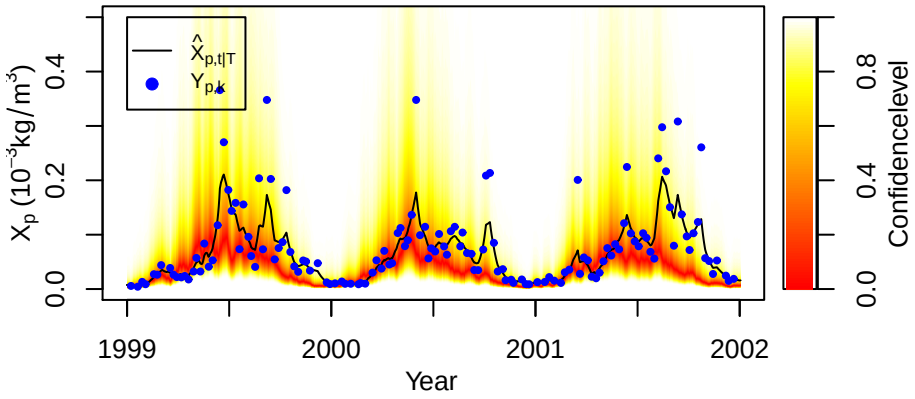

Figure 4: Time series of stationary distributions, smoothed state (black line) and observation (blue dots). The colour key refers to confidence intervals around the mode of the stationary distributions.

data. The higher flexibility of the estimating procedure, compared to discrete-time models, is a trade-off with the computational effort. The filter equations are computationally expensive and require many iterations when the number of parameters to be estimated is high. In practice, this limits the complexity of ecosystem models formulated as SDEs that can be estimated, acknowledging that the more complex mechanisms are contained in the stochastic processes of the covariance model and need regular updating through the filter equations. Consequently, SDEs are per se data-driven and less appropriate for long-term predictions or interpolation over larger gaps in the time series than ODEs. However, an important feature of SDEs is the uncertainty quantification of the model outputs, such uncertainty quantification cannot be readily and reliably provided by ODEs.

The real-data example provides results that perform well in the sense that the stationary distribution is close to data. It is clear, however, that extreme points are not well captured by the stationary distribution, and this effect could potentially be solved by including more effects (e.g. phytoplankton state and local weather conditions) in the growth process. Moreover, for the model to constitute a realistic representation with desired mathematical properties (stationarity of the solution), a two-state system, where phytoplankton remove nitrogen from the water column, is more appropriate. However, the aim of this study was to introduce SDEs and the estimation procedure, and not a modelling exercise. The more complex and hence adequate models will be considered in future research

studies.

References

Aït-Sahalia Y. (2008) Closed-form likelihood expansions for multivariate diffusions *The Annals of Statistics*, **32**, (2), 906-937

Baadsgaard M., Nielsen J. N., Spliid H., Madsen H., and Preisel M. (1997) Estimation in stochastic differential equations with state dependent diffusion term *SYSID '97 - 11th IFAC symposium on system identification, IFAC*

Bartell S.M., Lefebvre G., Kaminski G., Carreau M., and Campbell K. R. (1999) An ecosystem model for assessing ecological risks in Québec river, lakes and reservoirs *Ecological Modelling*, **124**, 43-67

Brillinger D.R., Preisler H.K., Ager A.A., Kie J.G., and Stewart B.S. (2002) Employing stochastic differential equation to model wildlife motion *Bulletin Brazilian Mathematical Society, New Series* , **33**(3), 385-408

Carpenter S.R., and Brock W.A. (2006) Rising variance: a leading indicator of ecological transition. *Ecology Letters*, **9**, 311-318

Dowd M. (2006) A sequential Monte Carlo approach for Marine ecological prediction *Environmetrics*, **17**, 435-455

Dowd M. (2006) Baysian statistical data assimilation for ecosystem models using Markov Chain Monte Carlo *Journal of Marine Systems*, **68**, 439-456

Fasham M.J.R, Ducklow H.W., and McKelvie S.M. (1990) A nitrogen-based model of plankton dynamics in the oceanic mixed layer *Journal of Marine Research*, **58**, 591-639

Givon D., Stinis P., and Weare J. (2009) Variance reduction for Particle filters of Systems With Time Scale Separation. *IEEE Transactions on Signal Processing*, **57**, (2), 424-435

Guttal V., and Jayaprakash C. (2008). Changing skewness: an early warning signal of regime shift in ecosystems. *Ecology Letters*, **201**(3), 420-428

Iacus S.M. (2008) Simulation and Inference for Stochastic Differential Equations - With R Examples *Springer Series in Statistics*

Jazwinski A.H. (1970) Stochastic Processes and Filtering Theory *Academic Press, New York, ASU*

Jacobsen J.L., and Madsen H. (1996). Grey Box Modelling of Oxygen Levels in a Small Stream, *Environmetrics* **7**, 109-121

Klebaner F.C. (2005) Introduction to Stochastic Calculus with Applications, *Imperial College Press*

Kristensen N.R., Madsen H., and Jørgensen S.B. (2004) Parameter estimation in stochastic grey-box models, *Automatica*, **40**, 225-237

Kristensen N.R., and Madsen H. (2003) Continuous time stochastic modeling - CTSM 2.3 - Mathematics Guide, *Technical University of Denmark*

Madsen H., Holst J., and Thyregod P. (1987). A Continuous Time Model for the Variations of Air Temperature, *10. Conference on Probability and Statistics in Atmospheric Science, American Meteorological Society,* 52-58, Edmonton

Madsen H., and Holst J. (1995) Estimation of Continuous-Time Models for the Heat Dynamics of a Building *Energy and Building,* **22**, 67-79

Limpert E., Stahel W.A. and Abbt M. (2001) Log-normal Distributions across the Sciences: Keys and Clues, *BioScience*, **51**(5),341-352

Luschgy H., and Pagés G. (2006) Functional quantization of a class of Brownian diffusions: A constructive approach *Stochastic Processes and their Applications* **116***, 310-336*

Madsen H. (2008) Time Series Analasis *Chapman & Hall/CRC*

Matear R. (1995) Parameter optimization and analasis of ecosystem models using simulated annealing: A case study at Station P. *Journal of Marine Research* **53**, 571-607

Nicolau, J. 2002. A new technique for simulating the likelihood of stochastic differential equations. *Econometrics* **5**, 91-103

Pastorello S., and Rossi E. (2010). Efficient importance sampling maximum likelihood estimation of stochastic differential equation, *Computational Statistics and Data Analasis* **54**, 2753-2762

Pedersen M. W., Righton D., Thygesen U.H., Andersen K., and H. Madsen. (2008). Geolocation of North Sea cod using Hidden Markov Models and behavioural switching, *Canadian Journal of Fisheries and Aquatic Sciences,* **65**, 2367-2377

Stollenwerk N., Friedhelm D.R., and Siegel H. Testing nonlinear stochastic models on phytoplankton biomass time series., *Ecological Modelling* **144**, 1261-277

Tornøe C.W., Jacobsen J., Pedersen O., Hansen T., and Madsen H. (2004). Grey-box Modelling of Pharmacokinetic/Pharmacodynamic Systems, *J. Pharmacokinet. Pharmacodyn,* **31**, 401-417

Øksendal B. (2003) Stochastic Differential Equations - An Introduction with Applications, Sixth edition. *Springer*, Berlin.

Paper D

From State Dependent Diffusion to Constant Diffusion in Stochastic Differential Equations by the Lamperti Transform

---

Authors:
J.K. Møller, and H. Madsen.

Published in:
IMM-technical report 2010-16, (2010).

Remarks:
The report is intended as a lecture note for the course in Advanced Time Series Analasis at DTU-Informatics.

From State Dependent Diffusion to Constant Diffusion in Stochastic Differential Equations by the Lamperti Transform

Jan Kloppenborg Møller[1,2], and Henrik Madsen[1]

Abstact

This report describes methods to eliminate state dependent diffusion terms in Stochastic Differential Equations (SDEs). Transformations that leave the diffusion term of SDEs constant is important for simulation, and estimation. It is important for simulation because the Euler approximation convergence rate is faster, and for estimation because the Extended Kalman Filter equations are easier to implement than higher order filters needed in the case of state dependent diffusion terms. The general class of transformations which leaves the diffusion term independent of the state is called the Lamperti transform. This note gives an example driven introduction to the Lamperti transform. The general applicability of the Lamperti transform is limited to univariate diffusion processes, but for a restricted class of multivariate diffusion processes Lamperti type transformations are available and the Lamperti transformation is discussed for both univariate and multivariate diffusion processes. Further some special attention is needed for time-inhomogeneous diffusion processes and these are discussed separately.

Key Words: Stochastic Differential Equations, Level dependent diffusion, Lamperti transform, Extended Kalman Filter.

1 Introduction

Stochastic differential equations (SDE's) are attracting increasing attention, because physical processes in real life systems experience random forcing, due to

[1]DTU Informatics, Richard Pedersens Plads, Technical University of Denmark Building 321, DK-2800 Lyngby, Denmark.

[2]National Environmental Research Institute, Fredriksborgvej 399, DK-4000 Roskilde, Denmark.

model approximations and stochastic inputs, that cannot be captured by ordinary differential equations (ODE's). Such random forcing or internal noise can be captured by adding random noise in the ODE, and this leads to SDE formulations.

The formulation of SDE's is done by physical reasoning. This physical reasoning includes autocorrelation structures and physical constraints (such as mass balance considerations) captured by the diffusion term. The formulation and reasoning often results in structures where the noise (diffusion) term depends on one or more state variables. Structures where the diffusion term depend on the state of the system are difficult to handle in estimation procedure like the one implemented in CTSM[1] (Kristensen & Madsen, 2003; Kristensen et. al., 2004), since the Extended Kalman Filter (EKF) requires higher (than 1) order terms in order to make the filter approximations sufficiently accurate. Therefore transformations that can move (or remove) the state dependence from the diffusion term to the drift term are needed. Other estimation procedures (Iacus, 2008) also rely on the existence of transformations of this sort. Transformations to unit diffusion is often referred to as Lamperti transform.

Further it is often recommended (Iacus, 2008) to use the Lamperti transformation before simulations. State dependent diffusion can together with structures in the drift term impose restrictions on the state space, e.g. processes that exist on the positive real axis only, like the Black and Scholes model (geometric Brownian motion). Estimation of such systems is not numerically stable if combined with a observation equation that use these constraints (like the log-transform), since estimation of the process may be zero (the geometric Brownian motion is strictly positive). However, after an appropriate transformation this process lives on the entire real axis and numerical problems on the boundary of the domain is avoided.

The results presented here seems to be well-known in more theoretical literature on SDE's (e.g. Luschgy, 2006), it is however hard to find papers, that explicitly deals with the construction of these kind of transformation in more applied settings. An exception is Nielsen & Madsen (2001), but comparing the results presented in that reference and the results presented here shows that the results in Nielsen & Madsen (2001) need corrections. Aït-Sahalia (2008) present transformations for a more general class of SDEs (referred to as reducible), these transformations are however more complicated to apply and we lose the generic formulations obtained in this report.

The report starts with a presentation of the general setting in Section 2. Results on one dimensional diffusion are given in Section 3, which is further divided

[1] `www.imm.dtu.dk/ctsm`

into time independent (Section 3.1) and time dependent diffusion (Section 3.2). The theoretical properties do not differ much between the two cases, but for practical applications some notes are needed for the time dependent diffusion. The multivariate case is presented in Section 4. This part does not consider a split into time independent and time dependent diffusion, since the remarks on the one dimensional time dependent case applies equally to the multidimensional case. Finally Section 5 gives a short summary and discussion of the result presented.

2 The general setting

Itô processes (SDE's) which are partly observed in discrete time are referred to as the continuous-discrete time stochastic state space models (Jazwinski, 1970), and a general formulation is

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta})d\boldsymbol{w}_t \tag{1}$$

$$\boldsymbol{Y}_k = \boldsymbol{g}(\boldsymbol{X}_{t_k}, t_k, \boldsymbol{u}_{t_k}, \boldsymbol{\theta}, \boldsymbol{e}_k), \tag{2}$$

where $t \in \mathbb{R}_0$ is time, $\boldsymbol{w}_t \in \mathbb{R}^m$ is the standard Brownian motion, $\boldsymbol{X}_t \in \mathbb{R}^n$ is the state variable, $\boldsymbol{u}_t \in \mathbb{R}^q$ is the input, $\boldsymbol{\theta} \in \mathbb{R}^p$ is a parameter vector, $\boldsymbol{f}(\cdot) \in \mathbb{R}^n$ is a vector function and $\boldsymbol{\sigma}(\cdot) \in \mathbb{R}^{n\times m}$ is the diffusion matrix. In the observation equation (2) $\boldsymbol{y} \in \mathbb{R}^l$ is the observations of state variable, $\boldsymbol{g} \in \mathbb{R}^l$ is the observation function and $\boldsymbol{e}_k \in \mathbb{R}^r$ is the observation error. The estimation problem is: Find $\hat{\boldsymbol{\theta}}$ such that

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}(S(\boldsymbol{\theta}, \mathcal{Y}_N)), \tag{3}$$

where S is some objective function and $\mathcal{Y}_N = \{\boldsymbol{Y}_1, ..., \boldsymbol{Y}_N\}$ is the set of all observations. The obvious choice for S is to maximise the one-step transitions probabilities, i.e. the product of the probability density functions (*pdf*'s) $p(Y_k|\mathcal{Y}_{k-1})$. This product is called the likelihood function (in practice we optimise the log-likelihood). The likelihood can in principle be found by solving the Fokker-Planck equation (Gard, 1988; Klebaner, 2005) and using Bayes rule for updating. It is however unrealistic to solve the Fokker-Planck equation if the system equation (1) does not have a very simple form. The general situation is sketched in Figure 1, to obtain the transition probability we need to integrate the SDE Eq. (1) between observations, when an observation is available the information provided by this observation is used to form the reconstruction of the state, and the transition probability to the next observation is again obtained by integration.

One way to move forward is by approximating the transition probabilities by Gaussian *pdf*'s, and transforming the observation equation such that the observation noise is (approximately) additive Gaussian. In order to calculate the likelihood function Extended Kalman Filters (EKFs) are often used, where the filter equations take complicated forms (higher order moments is needed and numerical solutions tend to be unstable if the diffusion term is a function of the state). It is therefore advisable to use transformations of the system equation (1) such that the diffusion is independent of the state. The transformation (ψ in (Figure 1) should form an equivalent relation between the input $\boldsymbol{u}_t$ and the output $\boldsymbol{Y}_k$ and the transformed system equation should depend on the same parameters ($\boldsymbol{theta}$). Even if the main problem is estimation, the application is more general since it is well-known that simulations has better convergence rates (Iacus, 2008) if the system equation is independent of the states. The subject of this note is transformations of the system equations that leaves the diffusion of the transformed system equations independent of the state.

In the following we will restrict the analysis to $\boldsymbol{\sigma}(\cdot) \in \mathbb{R}^{n\times n}$. There are two remarks about this 1) most derivations (except transformation to unity) generalise easily to the general case, and 2) in a weak solution sense (equality in distribution) this is not a restriction, since $\boldsymbol{\sigma}(\cdot)$ is only unique up to the (definite) "square root" of $\boldsymbol{\sigma}(\cdot)\boldsymbol{\sigma}^T(\cdot)$. A small example can illustrate the last point.

Example 1 *Consider the SDE*

$$dX_t = adt + \sigma_1 dw_{1,t} + \sigma_2 dw_{2,t}; \quad X_0 = 0, \tag{4}$$

where a, σ_1 and σ_2 are real constants. The solution to (4) is

$$X_t = at + \sigma_1 w_{1,t} + \sigma_2 w_{2,t}, \tag{5}$$

which is a Gaussian distributed random variable with mean and variance equal to at and $\sigma_1^2 + \sigma_2^2$, respectively, but this is also the (weak) solution to

$$dX_t = adt + \sqrt{\sigma_1^2 + \sigma_2^2} dw_t; \quad X_0 = 0, \tag{6}$$

which illustrates that the uniqueness of the weak solution is only unique up to the square root of $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$. □

The implication will be discussed further for multivariate processes in Section 4. The term "weak solution" refer to equality in distribution, and strong solutions refer to path-wise equality (see Øksendal (2003) for further discussions on weak and strong solutions). Clearly a strong solution is a weak solution, but a weak solution is not necessarily a strong solution (just consider Example 1). In this

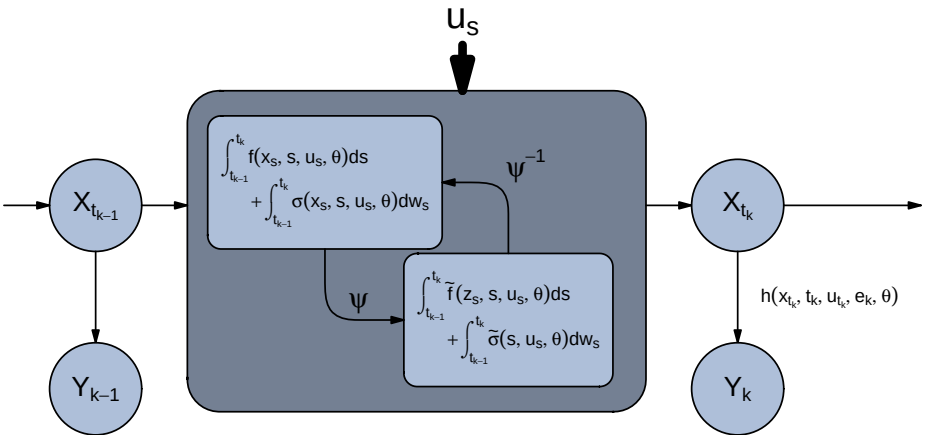

Figure 1: Conceptual diagram of the estimation problem, when an observation Y_k is available the state estimate of $X_{t_{k-1}}$ is updated by the provided information and used for integration of the state to form the prediction of the state X_{t_k}. There is an infinite number of equivalent relations between the input u_t and the output Y_k, the equivalence relation ψ gives a description with the same parameter, but $\tilde{\sigma}$ is independent of z_t.

note we will refer to weak solutions (which might also be strong) as solutions. In likelihood estimation the only interest is weak solutions, since we optimise the distribution. In simulation studies tha main interest will often also be weak solution.

2.1 Notation and problem setting

This note is only concerned with the system equation and with the comments above the class of differential equations is restricted to

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta})d\boldsymbol{w}_t, \tag{7}$$

where $\boldsymbol{\sigma} \in \mathbb{R}^{n\times n}$, $\boldsymbol{w}_t \in \mathbb{R}^n$ is the standard Brownian motion, and all other variable and functions are as explained below Eq. (1). This note deals with the problem; Find transformations $\boldsymbol{Z}_t = \boldsymbol{\psi}(\boldsymbol{X}_t, t)$ or $\tilde{\boldsymbol{Z}}_t = \tilde{\boldsymbol{\psi}}(\boldsymbol{X}_t, t)$ such that

$$d\boldsymbol{Z}_t = \widetilde{\boldsymbol{f}}(\boldsymbol{Z}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta})dt + \widetilde{\boldsymbol{\sigma}}(t, \boldsymbol{u}_t, \boldsymbol{\theta})d\boldsymbol{w}_t \tag{8}$$

or

$$d\tilde{\boldsymbol{Z}}_t = \widetilde{\boldsymbol{f}}_{\tilde{Z}}(\tilde{\boldsymbol{Z}}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta})dt + d\boldsymbol{w}_t, \tag{9}$$

where $\widetilde{\boldsymbol{\sigma}}(\cdot)$ is independent of $\boldsymbol{Z}_t$, but the parameters of (8) adn (9) are the same as in (7).

For notational convenience we will suppress the dependence of $\boldsymbol{\theta}$ and $\boldsymbol{u}_t$, i.e. we will use the notation

$$\boldsymbol{f}(\boldsymbol{X}_t, t) = \boldsymbol{f}(\boldsymbol{X}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta}) \tag{10}$$
$$\boldsymbol{\sigma}(\boldsymbol{X}_t, t) = \boldsymbol{\sigma}(\boldsymbol{X}_t, t, \boldsymbol{u}_t, \boldsymbol{\theta}). \tag{11}$$

In real life systems $\boldsymbol{u}_t$ is often a set of observations, i.e. not a function that can be differentiated analytically, and this has to be kept in mind in the following development of the transformations.

3 One dimensional diffusion

The fundamental tool for transformations of SDE's is Itô's lemma (the version given below is due to Øksendal (2003))

Theorem 1 (Itô's lemma): *Let X_t be an Itô process given by*

$$dX_t = f(X_t, t)dt + \sigma(X_t, t)dw_t. \tag{12}$$

Let $\psi(X_t, t) \in C^2([0, \infty) \times \mathbb{R}$. Then

$$Z_t = \psi(X_t, t) \tag{13}$$

is again an Itô process, and

$$dZ_t = \frac{\partial \psi}{\partial t}(X_t, t)dt + \frac{\partial \psi}{\partial x}(X_t, t)dX_t + \frac{1}{2}\frac{\partial^2 \psi}{\partial x^2}(X_t, t)(dX_t)^2, \tag{14}$$

where $(dX_t)^2$ is calculated according to the rules

$$dt \cdot dt = dt \cdot dw_t = dw_t \cdot dt = 0, \quad dw_t \cdot dw_t = dt. \tag{15}$$

The proof of this theorem is out of the scope of this note, and the reader is referred to Øksendal (2003).

It is illustrative to express Itô's formula in terms of dw_t rather than dX_t. For notational reasons we will sometimes write f for $f(X_t, t)$, σ for $\sigma(X_t, t)$ and

ψ for $\psi(X_t, t)$, partial derivatives will be written as $\psi_s = \frac{\partial \psi}{\partial s}$ and $\psi_{ss} = \frac{\partial^2 \psi}{\partial s^2}$. Rearranging (14) gives

$$dZ_t = \psi_t dt + \psi_x \cdot (f dt + \sigma dw_t) + \frac{1}{2}\psi_{xx} \cdot (f dt + \sigma dw_t)^2 \tag{16}$$

$$= (\psi_t + \psi_x \cdot f)\, dt + \psi_x \cdot \sigma dw_t + \frac{1}{2}\psi_{xx} \cdot \sigma^2 dt \tag{17}$$

$$= \left(\psi_t + \psi_x \cdot f + \frac{1}{2}\psi_{xx} \cdot \sigma^2\right) dt + \psi_x \cdot \sigma dw_t. \tag{18}$$

With this formulation we are ready for the construction of a transformation for removal of level dependent noise. The following constructive theorem is often referred to as the Lamperti transform (Iacus, 2008; Luschgy, 2006).

Theorem 2 (Lamperti transform): *Let X_t be an Itô process as in (12), and define*

$$\psi(X_t, t) = \int \frac{1}{\sigma(x,t)} dx \Bigg|_{x=X_t}, \tag{19}$$

if ψ is one to one from the state space of X_t onto $\mathbb{R}$ for every $t \in [0, \infty)$, then choose $Z_t = \psi(X_t, t)$. Otherwise if $\sigma(X_t, t) > 0 \quad \forall (X_t, t)$ choose

$$Z_t = \psi(X_t, t) = \int_\xi^x \frac{1}{\sigma(u,t)} du \Bigg|_{x=X_t}, \tag{20}$$

where ξ is some point inside the state space of X_t. Then Z_t has unit diffusion and is governed by the SDE

$$dZ_t = \left(\psi_t(\psi^{-1}(Z_t, t), t) + \frac{f(\psi^{-1}(Z_t, t), t)}{\sigma(\psi^{-1}(Z_t, t), t)} - \frac{1}{2}\sigma_x(\psi^{-1}(Z_t, t), t)\right) dt + dw_t. \tag{21}$$

A transformation of the state-space clearly has to be one to one, such that every point in the state space of X_t can be uniquely identified by the inverse transformation of Z_t. If (19) is not one to one, then choosing the transformation (20) (due to Luschgy (2006)) will ensure that the transformation is one to one, since ψ is then a strictly increasing function of X_t. We will prove Eq. (19) and leave Eq. (20) to the reader.

Proof. (Of Theorem 2) From (18) it is easy to realize that level dependent diffusion can be removed by choosing the transformation ψ as

$$\psi(X_t, t) = \int \frac{1}{\sigma(x,t)} dx \Bigg|_{x=X_t} \quad \Longrightarrow \quad \psi_x(X_t, t) = \frac{1}{\sigma(X_t, t)}. \tag{22}$$

Differentiation w.r.t. x and time gives

$$\psi_{xx}(X_t,t) = -\frac{\sigma_x(X_t,t)}{\sigma(X_t,t)^2} \tag{23}$$

$$\psi_t(X_t,t) = \frac{\partial}{\partial t}\int \frac{1}{\sigma(x,t)}dx\bigg|_{x=X_t}. \tag{24}$$

Inserting in (18) gives

$$dZ_t = \left(\frac{\partial}{\partial t}\int \frac{1}{\sigma(x,t)}dx\bigg|_{x=X_t} + \frac{f}{\sigma} - \frac{1}{2}\frac{\sigma_x}{\sigma^2}\sigma^2\right)dt + dw_t. \tag{25}$$

Cancelling out denominators and enumerators and inserting ψ_t and $X_t = \psi^{-1}(Z_t,t)$ gives the desired result. □

Theorem 2 gives a very useful approach for removal of level dependent noise. The discussion of the theorem in the following, is largely example driven and divided in two parts. 1) Time independent diffusion i.e. $\psi_t = 0$, and 2) time dependent diffusion.

3.1 Time independent diffusion

We begin this section with a small example, which illustrates the use of the Lamperti transformation.

Example 2 (Geometric Brownian motion): *Let X_t be an Itô process (SDE) given by*

$$dX_t = aX_t dt + \sigma X_t dw_t; \quad X_0 = 1, \tag{26}$$

where σ and a are real constants. Choose ψ as in (19), i.e.

$$Z_t = \psi(X_t) = \int \frac{1}{\sigma x}dx\bigg|_{x=X_t} = \frac{\log(X_t)}{\sigma}, \tag{27}$$

and

$$X_t = \psi^{-1}(Z_t) = e^{\sigma Z_t}. \tag{28}$$

By (21) Z_t is an Itô process given by

$$dZ_t = \left(\frac{aX_t}{\sigma X_t} - \frac{1}{2}\sigma\right) dt + dw_t \tag{29}$$

$$= \left(\frac{a}{\sigma} - \frac{1}{2}\sigma\right) dt + dw_t. \tag{30}$$

In this case the solution of Z_t can be given explicitly as

$$Z_t = \left(\frac{a}{\sigma} - \frac{1}{2}\sigma\right) t + w_t, \tag{31}$$

i.e. $Z_t \sim N\left(\left(\frac{a}{\sigma} - \frac{1}{2}\sigma\right) t, t\right) \Rightarrow \sigma Z_t \sim N\left(\left(a - \frac{1}{2}\sigma^2\right) t, \sigma^2 t\right)$, since $X_t = e^{\sigma Z_t}$, the solution of X_t is given as $X_t \sim LN\left(\left(a - \frac{1}{2}\sigma^2\right) t, \sigma^2 t\right)$, where LN is the log-normal distribution. □

In the example above the Lamperti transform actually solves the original equation. This is not the case in general, but Itô's formula can be used to solve SDE's, although the class of equations that are solvable in this fashion is limited. The inverse transform of Z_t was not a part of the SDE governing Z_t, this is not the case in general, and SDE's that are apparently very simple cannot be solve explicitly, as the next example illustrates.

Example 3 *Consider*

$$dX_t = (b + aX_t)dt + \sigma X_t dw_t. \tag{32}$$

Using (19) we get the same transformation as in Example 2, and Z_t is governed by

$$dZ_t = \left(\frac{b + aX_t}{\sigma X_t} - \frac{1}{2}\sigma\right) dt + dw_t \tag{33}$$

$$= \left(\frac{b}{\sigma} e^{-\sigma Z_t} + \frac{a}{\sigma} - \frac{1}{2}\sigma\right) dt + dw_t, \tag{34}$$

this SDE does not have an explicit solution, but parameter estimation is available through e.g. CTSM and numerical solutions (i.e. the distribution) can be found through simulations.

Eq. (19) is in principle always valid. The practical application of the transformation is, however, limited by our ability of find an explicit solution of the inverse transformation

$$X_t = \psi^{-1}(Z_t, t). \tag{35}$$

Such solutions are not always available as illustrated in the following example

Example 4 *Consider the diffusion process*

$$dX_t = f(X_t)dt + (\sigma_0 + \sigma_1\sqrt{X_t})dw_t. \tag{36}$$

The Lamperti transform becomes

$$Z_t = \psi(X_t) = \frac{\sigma_0}{\sigma_1^2}\log\left((\sigma_0 + \sigma_1\sqrt{X_t})^{-2}\right) + \frac{2}{\sigma_1}\sqrt{X_t} \tag{37}$$

$$= \frac{2}{\sigma_1}\left(\sqrt{X_t} - \frac{\sigma_0}{\sigma_1}\log\left(\sigma_0 + \sigma_1\sqrt{X_t}\right)\right). \tag{38}$$

In this case the Itô diffusion of Z_t cannot be written as an explicit function of Z_t, because X_t cannot be written as an explicit function of Z_t. □

As illustrated by Example 4 explicit solutions for the inverse transformation does not always exist, however many "real" life examples allow the explicit solution of the inverse transform. For instance explicit solutions of ψ^{-1} is available when $\sigma(X_t) = \sigma_1 X_t^\gamma$ for any constant γ, models of this type important in mathematical finance, where γ express the volatility of the market.

For models where $\sigma(X_t)$ are more complex, solutions to ψ^{-1} are in general not available. Biological models often use proportional or square root dependent diffusion terms, and in addition additive diffusion might be appropriate if the model contain additive input. As we saw in Example 4, ψ^{-1} is not available in this case. A quite flexible system where ψ^{-1} is available is the Pearson diffusion (Forman and Sørensen, 2008), which is considered in the following example.

Example 5 (Pearson diffusion): *Consider the diffusion process*

$$dX_t = f(X_t)dt + \sqrt{\sigma_0 + \sigma_1 X_t + \sigma_2 X_t^2}dw_t. \tag{39}$$

Actually this is an extension of the Pearson diffusion as the Pearson diffusion also have $f(X_t) = (b - aX_t)$. In this context we will, however, only consider the diffusion term and assume $\sigma_i > 0$. Use of the Lamperti transform (19) gives

$$Z_t = \psi(X_t) = \frac{1}{\sqrt{\sigma_2}}\log\left(\frac{\sigma_1}{2\sqrt{\sigma_2}} + \sqrt{\sigma_2}X_t + \sqrt{\sigma_0 + \sigma_1 X_t + \sigma_2 X_t^2}\right) \tag{40}$$

with the inverse

$$X_t = \psi^{-1}(Z_t) = \left(\frac{\sigma_1^2}{8\sigma_2^{3/2}} - \frac{\sigma_0}{2\sqrt{\sigma_2}}\right)e^{-\sqrt{\sigma_2}Z_t} + \frac{1}{2\sqrt{\sigma_2}}e^{\sqrt{\sigma_2}Z_t} - \frac{\sigma_1}{2\sigma_2} \tag{41}$$

and the Itô process for Z_t is given by

$$dZ_t = \left(\frac{f(\psi^{-1}(Z_t))}{\sqrt{\sigma_0 + \sigma_1\psi^{-1}(Z_t) + \sigma_2 \cdot (\psi^{-1}(Z_t))^2}} - \frac{\sigma_1 + 2\sigma_2\psi^{-1}(Z_t)}{4\sqrt{\sigma_0 + \sigma_1\psi^{-1}(Z_t) + \sigma_2 \cdot (\psi^{-1}(Z_t))^2}} \right) dt + dw_t \tag{42}$$

$$= \frac{f(\psi^{-1}(Z_t)) - \frac{1}{4}\left(\sigma_1 + 2\sigma_2\psi^{-1}(Z_t)\right)}{\sqrt{\sigma_0 + \sigma_1\psi^{-1}(Z_t) + \sigma_2 \cdot (\psi^{-1}(Z_t))^2}} dt + dw_t. \tag{43}$$

Clearly the resulting SDE is very complex, it will however provide the opportunity of testing hypothesis of $\sigma_i = 0$. In the construction of SDE's of the type discussed in this example it is important to ensure that the diffusion term exists for all X_t in the state space of X_t (we would need to examine the drift term at the boundary). □

Even though the Lamperti transform is limited by our ability of finding the inverse, it is still possible to use transformations that remove level dependent noise for quite general classes of diffusion processes, as illustrated in Example 5.

3.2 Time dependent diffusion

The SDE (21) depends on the time derivative of ψ, and even though this might be a quite complicated function, it is in principle always possible to find such a solution. In real life applications the time dependence of σ will, however, often be through an observed input, in this case the differentiation have to be done numerically. It might therefore be advisable to choose a transformation that leaves the diffusion term time dependent. This does however limit the the class of transformations substantially, it is e.g. not possible if one of the diffusion parameters in the Pearson diffusion depends on time.

In general it is possible to succeed in the case where the diffusion is given by

$$\sigma(X_t, t) = \alpha(t)\beta(X_t) \tag{44}$$

In this case use the Lamperti transform on $\beta(X_t)$ and leave the diffusion time-dependent, i.e. put

$$Z_t = \psi(X_t) = \int \frac{1}{\beta(x)} dx \Bigg|_{x=X_t}, \tag{45}$$

and proceeding like in the time-independent diffusion we get

$$dZ_t = \left(\frac{f(\psi^{-1}(Z_t), t)}{\beta(\psi^{-1}(Z_t))} - \frac{1}{2}\beta_x(\psi^{-1}(Z_t))\alpha^2(t) \right) dt + \alpha(t)dw_t. \tag{46}$$

If the time dependence is either an explicit function of t or the differential of the time dependence is available through observations then Theorem 2 can still be applied, but the functional relationships do however become considerable more complex, as the next example illustrates.

Example 6 *Consider a process driven by a noisy time varying input $b(t)$ (birth process) and with a constant death-rate, the SDE formulation could be*

$$dX_t = (b(t) + aX_t)dt + (\sigma_0 b(t) + \sigma_1 X_t)dw_t, \tag{47}$$

where $a > 0$ and $b(t) > 0 \quad \forall t$. The Lamperti transform becomes

$$\psi(X_t, t) = \int \frac{1}{\sigma_0 b(t) + \sigma_1 x} dx \bigg|_{x=X_t} = \frac{\log(\sigma_0 b(t) + \sigma_1 x)}{\sigma_1} \tag{48}$$

implying

$$\psi_t(X_t, t) = \frac{\sigma_0 b'(t)}{\sigma_1(\sigma_0 b(t) + \sigma_1 X_t)} \tag{49}$$

$$\psi^{-1}(Z_t, t) = \frac{e^{\sigma_1 Z_t} - \sigma_0 b(t)}{\sigma_1} \tag{50}$$

$$\sigma_x(X_t, t) = \sigma_1, \tag{51}$$

and $Z_t = \psi(X_t, t)$ is governed by the process

$$dZ_t = \left(\frac{\sigma_0 b'(t)}{\sigma_1(\sigma_0 b(t) + \sigma_1 \psi^{-1}(Z_t, t))} + \frac{b(t) + a\psi^{-1}(Z_t, t)}{\sigma_0 b(t) + \sigma_1 \psi^{-1}(Z_t, t)} - \frac{1}{2}\sigma_1 \right) dt + dw_t \tag{52}$$

$$= \left(\frac{\frac{\sigma_0}{\sigma_1} b'(t) + b(t) + a\frac{e^{\sigma_1 Z_t} - \sigma_0 b(t)}{\sigma_1}}{\sigma_0 b(t) + \sigma_1 \frac{e^{\sigma_1 Z_t} - \sigma_0 b(t)}{\sigma_1}} - \frac{1}{2}\sigma_1 \right) dt + dw_t \tag{53}$$

$$= \left[\left(\frac{\sigma_0}{\sigma_1} b'(t) + \left(1 - a\frac{\sigma_0}{\sigma_1}\right) b(t) + \frac{a}{\sigma_1} e^{\sigma_1 Z_t} \right) e^{-\sigma_1 Z_t} - \frac{1}{2}\sigma_1 \right) dt + dw_t$$

$$= \left\{ \left[\frac{\sigma_0}{\sigma_1} b'(t) + \left(1 - a\frac{\sigma_0}{\sigma_1}\right) b(t) \right] e^{-\sigma_1 Z_t} + \frac{a}{\sigma_1} - \frac{1}{2}\sigma_1 \right\} dt + dw_t. \tag{54}$$

In principle this is straight forward, but $b(t)$ will often be a function of some observed process and in this case we will therefore need observations of the differential of $b(t)$.

4 Multivariate Diffusion

The Lamperti transform presented so far is a univariate transformation, but it is possible to generalise this for a restricted class of multivariate diffusion processes. As for the one-dimensional diffusion process, Itô's lemma for multi-dimensional diffusion is the key to understand the multi-dimensional transformation. Again a good reference is Øksendal (2003).

Theorem 3 (Itô's lemma):

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, t)dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, t,)d\boldsymbol{w}_t, \tag{55}$$

with $t \in \mathbb{R}_+$ *being time,* $\boldsymbol{X}_t \in \mathbb{R}^n$ *the state vector,* $\boldsymbol{w}_t \in \mathbb{R}^n$ *multivariate standard Brownian motion,* $\boldsymbol{f}(\cdot) \in \mathbb{R}^n$ *and* $\boldsymbol{\sigma}(\cdot) \in \mathbb{R}^{n\times n}$*. Then for a given transformation*

$$\boldsymbol{Z}_t = \boldsymbol{\psi}(\boldsymbol{X}_t, t) = [\psi_1(\boldsymbol{X}_t, t), ..., \psi_n(\boldsymbol{X}_t, t)], \tag{56}$$

where $\boldsymbol{\psi}$ *is a* C^2 *function from* $\mathbb{R}^n \times [0, \infty)$ *into* $\mathbb{R}^n$*,* $\boldsymbol{Z}_t$ *is again an Itô process given by*

$$\begin{aligned} dZ_{k,t} =& \frac{\partial \psi_k}{\partial t}(\boldsymbol{X}_t, t)dt + \sum_{i=1}^{n} \frac{\partial \psi_k}{\partial x_i}(\boldsymbol{X}_t, t)dX_{i,t} + \\ & \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 \psi_k}{\partial x_i \partial x_j}(\boldsymbol{X}_t, t)dX_{j,t}dX_{i,t}. \end{aligned} \tag{57}$$

Where $dw_i dw_j = 0$ *for* $i \neq j$*,* $dw_i dw_i = dt$*, and* $dw_i dt = dt dw_i = dt dt = 0 \quad \forall i$*.*

In the version of Itô's lemma given above $\boldsymbol{\psi} \in \mathbb{R}^n$. In the general version of Itô's lemma this not a requirement, but we have restricted the attention to equal dimensions of $\boldsymbol{X}_t$ and $\boldsymbol{Z}_t$. The derivations below do however easily generalise.

It is again illustrative to write $dZ_{k,t}$ in terms of $dw_{i,t}$ rather than $dX_{i,t}$,

$$\begin{aligned} dZ_{k,t} =& \frac{\partial \psi_k}{\partial t}(\boldsymbol{X}_t, t)dt + \sum_{i=1}^{n} \frac{\partial \psi_k}{\partial x_i}(\boldsymbol{X}_t, t)dX_{i,t} + \\ & \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 \psi_k}{\partial x_i \partial x_j}(\boldsymbol{X}_t, t)dX_{j,t}dX_{i,t} \qquad (58) \\ =& \left((\psi_k)_t + \sum_{i=1}^{n} (\psi_k)_{x_i} f_i \right) dt + \sum_{i=1}^{n} (\psi_k)_{x_i} \left(\sum_{h=1}^{n} \sigma_{ih} dw_{h,t} \right) + \\ & \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} (\psi_k)_{x_i,x_j} \left(\sum_{h=1}^{n} \sigma_{jh} dw_{h,t} \right) \left(\sum_{l=1}^{m} \sigma_{jl} dw_{l,t} \right) \qquad (59) \\ =& \left((\psi_k)_t + \sum_{i=1}^{n} (\psi_k)_{x_i} f_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} (\psi_k)_{x_i,x_j} \left(\sum_{h=1}^{m} \sigma_{jh}\sigma_{ih} \right) \right) dt + \\ & \sum_{h=1}^{m} \left(\sum_{i=1}^{n} (\psi_k)_{x_i} \sigma_{ih} \right) dw_{h,t}, \qquad (60) \end{aligned}$$

where subscript $\{h, i, j, k\}$ refer to elements of vectors and matrices, subscripts x_i and t refer to partial differentiation (except in $Z_{i,t}$ and $w_{i,t}$ where t refer to time). From the last expression in Eq. (60) it is seen that the removal of level dependent noise requires the solution of the following system of PDEs

$$\begin{aligned} \sum_{i=1}^{n} (\psi_k)_{x_i} \sigma_{i1}(\boldsymbol{x}, t,) &= c_1(t) \qquad (61) \\ \sum_{i=1}^{n} (\psi_k)_{x_i} \sigma_{i2}(\boldsymbol{x}, t) &= c_2(t) \qquad (62) \\ &\vdots \\ \sum_{i=1}^{n} (\psi_k)_{x_i} \sigma_{in}(\boldsymbol{x}, t) &= c_n(t), \qquad (63) \end{aligned}$$

where c_i is an arbitrary function of t. Such a system can not be solved in general, since for given $\boldsymbol{\sigma}$, this results in n equations with one unknown (ψ_k).

Nielsen & Madsen (2001) claim that under the assumptions 1) $\sigma_{ij} \neq 0$ and 2)

$$\sigma_{ij}(\boldsymbol{X}_t, t) = \sigma_{ij}(X_t^{\nu(i)}, t), \quad i = 1, ..., n, j = 1, ..., n, \qquad (64)$$

it is possible to find a transformation. The application of Itô lemma is however wrong, we will not go through the proof of this, but applying (61)-(63) will lead to $\nu(i) = i$.

The difficulties of removing state dependent diffusion can be illustrated by a simple example.

Example 7 *Consider the diffusion process*

$$d\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} X_{2,t} & 0 \\ 0 & X_{1,t} \end{bmatrix} \begin{bmatrix} dw_{1,t} \\ dw_{2,t} \end{bmatrix}, \tag{65}$$

let $Z_{1,t} = \psi_1(\boldsymbol{X}_t)$ *and* $Z_{2,t} = \psi_2(\boldsymbol{X}_t)$*. Using Itô's lemma we get*

$$\begin{aligned} dZ_{1,t} =& \frac{\partial}{\partial x_1}\psi_1(X_{1,t}, X_{2,t})X_{2,t}dw_{1,t} + \frac{\partial}{\partial x_2}\psi_1(X_{1,t}, X_{2,t})X_{1,t}dw_{2,t} + \\ & \frac{1}{2}\left(\frac{\partial^2}{\partial x_1 \partial x_1}\psi_1(X_{1,t}, X_{2,t})X_{2,t}^2 + \frac{\partial^2}{\partial x_2 \partial x_2}\psi_1(X_{1,t}, X_{2,t})X_{1,t}^2\right)dt, \end{aligned} \tag{66}$$

the first term requires the solution of

$$c_1 = x_2 \frac{\partial}{\partial x_1}\psi_1(x_1, x_2) \tag{67}$$

implying

$$\psi_1(x_1, x_2) = c\frac{x_1}{x_2} + \tilde{\psi}_1(x_2). \tag{68}$$

where $\tilde{\psi}_1(x_2)$ *is an arbitrary function of* x_2*. The second term therefore require the solution of*

$$c_2 = \frac{\partial}{\partial x_2}\left(c_1\frac{x_1}{x_2} + \tilde{\psi}_1(x_2)\right) \tag{69}$$

$$= -c_1\frac{x_1}{x_2^2} + \frac{d}{dx_2}\tilde{\psi}_1(x_2), \tag{70}$$

as $\tilde{\psi}_1(x_2)$ *is not a function of* x_1 *this differential equation does not admit a solution.* □

Even if a general multivariate version of Theorem 2 is not available, it is possible to state the less general result

Theorem 4 (Multivariate Lamperti transform): *Let* X_t *be an Itô diffusion given by*

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, t)dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, t)\boldsymbol{R}(t)d\boldsymbol{w}_t, \tag{71}$$

where $\boldsymbol{R}(t) \in \mathbb{R}^{n\times n}$ *is any matrix function of* t*, and* $\boldsymbol{\sigma}(\boldsymbol{X}_t, t) \in \mathbb{R}^{n\times n}$ *is a diagonal matrix with diagonal elements* $\sigma_{i,i}(\boldsymbol{X}_t, t)$ *given by*

$$\sigma_{i,i}(\boldsymbol{X}_t, t) = \sigma_i(X_{i,t}, t). \tag{72}$$

Then the transformation

$$Z_{i,t} = \psi_i(X_{i,t}, t) = \int \frac{1}{\sigma_i(x,t)} dx \Bigg|_{x=X_{i,t}}, \tag{73}$$

will result in an Itô process with the i'th element given by

$$\begin{aligned} dZ_{i,t} = \Bigg(& \frac{\partial}{\partial t}\psi_i(x,t)\big|_{x=\psi^{-1}(Z_{i,t},t)} + \frac{f_i(\boldsymbol{\psi}^{-1}(\boldsymbol{Z}_t,t),t)}{\sigma_i(\psi_i^{-1}(Z_{i,t},t),t)} - \\ & \frac{1}{2}\frac{\partial}{\partial x}\sigma_i(\psi_i^{-1}(Z_{i,t},t),t) \cdot \sum_{j=1}^{n} r_{ij}^2(t) \Bigg) dt + \sum_{j=1}^{n} r_{ij}(t) dw_{j,t}, \end{aligned} \tag{74}$$

where $r_{ij}(t)$ *are elements of* $\boldsymbol{R}(t)$ *and*

$$\boldsymbol{X}_t = \boldsymbol{\psi}^{-1}(\boldsymbol{Z}_t, t). \tag{75}$$

Proof. Apply Theorem 2 to each $X_{i,t}$ □

The remarks about time dependent diffusion made in Section 3.2 also apply to the multidimensional case. A simple example illustrates the use of Theorem 4.

Example 8 (Two-dimensional Geometric Brownian motion): *Consider the process*

$$d\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} dt + \begin{bmatrix} X_{1,t} & 0 \\ 0 & X_{2,t} \end{bmatrix} \begin{bmatrix} r_{1,1} & r_{1,2} \\ r_{2,1} & r_{2,2} \end{bmatrix} d\boldsymbol{w}_t \tag{76}$$

$$= \boldsymbol{A}\boldsymbol{X}_t dt + \boldsymbol{\sigma}(\boldsymbol{X}_t)\boldsymbol{R} d\boldsymbol{w}_t, \tag{77}$$

with initial condition $\boldsymbol{X}_0 = 1$*, choose* $Z_{1,t} = \psi(\boldsymbol{X}_t) = \log(X_{1,t})$ *and* $Z_{2,t} = \psi(\boldsymbol{X}_t) = \log(X_{2,t})$*, then*

$$d\begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} = \begin{bmatrix} a_1 - \frac{1}{2}(r_{11}^2 + r_{12}^2) \\ a_2 - \frac{1}{2}(r_{21}^2 + r_{22}^2) \end{bmatrix} dt + \begin{bmatrix} r_{1,1} & r_{1,2} \\ r_{2,1} & r_{2,2} \end{bmatrix} d\boldsymbol{w}_t, \tag{78}$$

with initial condition $\boldsymbol{Z}_0 = 0$ *and the solution of* (78) *is*

$$\begin{bmatrix} Z_{1,t} \\ Z_{2,t} \end{bmatrix} = \begin{bmatrix} a_1 - \frac{1}{2}(r_{11}^2 + r_{12}^2) \\ a_2 - \frac{1}{2}(r_{21}^2 + r_{22}^2) \end{bmatrix} t + \begin{bmatrix} r_{1,1} & r_{1,2} \\ r_{2,1} & r_{2,2} \end{bmatrix} \boldsymbol{w}_t \tag{79}$$

$$= (\boldsymbol{a} - diag(\boldsymbol{R}\boldsymbol{R}^T))t + \boldsymbol{R}\boldsymbol{w}_t, \tag{80}$$

where $\boldsymbol{a} = diag(\boldsymbol{A})$*is a vector with elements equal to the diagonal elements of* $\boldsymbol{A}$*. In this case* $\boldsymbol{Z}_t$ *follows a Gaussian distribution* $\boldsymbol{Z}_t \sim N((\boldsymbol{a} - \frac{1}{2} diag(\boldsymbol{R}\boldsymbol{R}^T))t, \boldsymbol{R}\boldsymbol{R}^T t)$*.* $\boldsymbol{X}_t$ *is therefore distributed according to a multivariate* log*-normal distribution, with the same parameters.* □

The transformation presented in Theorem 4 is not a true Lamperti transform, since it does not transform to unit diffusion. This can be solved by the following theorem

Theorem 5 (Tranformation to unit diffusion): *Let* X_t *be an Itô diffusion given by*

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, t)dt + \boldsymbol{R}(t)d\boldsymbol{w}_t, \tag{81}$$

where $\boldsymbol{R}(t) \in \mathbb{R}^{n \times n}$ *is any invertible matrix function of* t*, then the transformation*

$$\boldsymbol{Z}_{i,t} = \boldsymbol{\psi}(\boldsymbol{X}_t, t) = \boldsymbol{R}(t)^{-1}\boldsymbol{X}_t \tag{82}$$

will result in an Itô process given by

$$d\boldsymbol{Z}_t = \left[\left(\frac{d}{dt}\boldsymbol{R}(t)^{-1}\right)\boldsymbol{R}(t)\boldsymbol{Z}_t + \boldsymbol{R}(t)^{-1}\boldsymbol{f}(\boldsymbol{R}(t)\boldsymbol{Z}_t, t)\right]dt + d\boldsymbol{w}_t \tag{83}$$

$$\tag{84}$$

with $(\frac{d}{dt}\boldsymbol{R}(t)^{-1})$ *the elements-wise derivative of* $\boldsymbol{R}(t)^{-1}$ *and*

$$\boldsymbol{X}_t = \boldsymbol{R}(t)\boldsymbol{Z}_t. \tag{85}$$

Proof. Consider the i'th coordinate of $d\boldsymbol{X}_t$

$$dX_{i,t} = f_i(\boldsymbol{X}_t, t)dt + \sum_{j=1}^{n}(\boldsymbol{R}(t))_{i,j}dw_j, \tag{86}$$

and the i'th coordinate of $\boldsymbol{Z}_t$

$$Z_{i,t} = \psi_i(\boldsymbol{X}_t, t) = \sum_{j=1}^{n}(\boldsymbol{R}(t)^{-1})_{ij}X_{j,t} \tag{87}$$

by Itô's lemma we get (noting that $\frac{\partial^2}{\partial x_i \partial x_j}\psi_i(\boldsymbol{x},t) = 0 \ \forall \ i \in \{1,..,n\}$, and $j \in \{1,...,n\}$)

$$
\begin{aligned}
Z_{i,t} =& \frac{\partial}{\partial t}\psi_i(\boldsymbol{X}_t,t)dt + \sum_{j=1}^{n}\frac{\partial}{\partial x_j}\psi_i(\boldsymbol{X}_t,t)dX_{j,t} && (88)\\
=& \frac{\partial}{\partial t}\sum_{j=1}^{n}(\boldsymbol{R}(t)^{-1})_{ij}X_{j,t}dt + \sum_{j=1}^{n}(\boldsymbol{R}(t)^{-1})_{ij}[f_j(\boldsymbol{X}_t,t)dt + \\
& \sum_{h=1}^{n}\boldsymbol{R}(t)_{jh}dw_h] && (89)\\
=& \left[\sum_{j=1}^{n}\frac{d}{dt}\boldsymbol{R}(t)_{ij}^{-1}X_{j,t} + \sum_{j=1}^{n}\boldsymbol{R}(t)_{ij}^{-1}f_j(\boldsymbol{X}_t,t)\right]dt + \\
& \sum_{j=1}^{n}\boldsymbol{R}(t)_{ij}^{-1}\sum_{h=1}^{n}\boldsymbol{R}(t)_{jh}dw_h && (90)\\
=& \left[\sum_{j=1}^{n}\left(\frac{d}{dt}\boldsymbol{R}(t)_{ij}^{-1}\right)\sum_{h=1}^{n}\boldsymbol{R}(t)_{jh}^{-1}Z_{h,t} + \left(\boldsymbol{R}(t)^{-1}\boldsymbol{f}(\boldsymbol{X}_t,t)\right)_i\right]dt + \\
& \sum_{h=1}^{n}\left(\sum_{j=1}^{n}\boldsymbol{R}(t)_{ij}^{-1}\boldsymbol{R}(t)_{jh}\right)dw_h && (91)\\
=& \left[\sum_{h=1}^{n}\left(\left(\frac{d}{dt}\boldsymbol{R}(t)^{-1}\right)\boldsymbol{R}(t)\right)_{ih}Z_{h,t} + \left(\boldsymbol{R}(t)^{-1}\boldsymbol{f}(\boldsymbol{X}_t,t)\right)_i\right]dt \\
& \sum_{h=1}^{n}\left(\boldsymbol{R}(t)^{-1}\boldsymbol{R}(t)\right)_{ih}dw_h && (92)\\
=& \left[\left(\left(\frac{d}{dt}\boldsymbol{R}(t)^{-1}\right)\boldsymbol{R}(t)\boldsymbol{Z}_t\right)_i + \left(\boldsymbol{R}(t)^{-1}\boldsymbol{f}(\boldsymbol{X}_t,t)\right)_i\right]dt + dw_i, && (93)
\end{aligned}
$$

writing the matrix formulation of the above gives the desired result. □

Combining Theorem 4 and 5 gives a multivariate version of the Lamperti transform. This is illustrated by applying Theorem 5 to Example 8.

Example 9 *Consider the transformed process of Example 8, then*

$$
\boldsymbol{R}^{-1} = \frac{1}{\det(\boldsymbol{R})}\begin{bmatrix} r_{22} & -r_{12} \\ -r_{21} & r_{11} \end{bmatrix}, \qquad (94)
$$

and the process

$$\tilde{\boldsymbol{Z}}_t = \tilde{\boldsymbol{\psi}}(\boldsymbol{Z}_t) = \boldsymbol{R}^{-1}\boldsymbol{Z}_t, \tag{95}$$

is given by

$$\begin{aligned} d\begin{bmatrix} \tilde{Z}_{1,t} \\ \tilde{Z}_{2,t} \end{bmatrix} =& \frac{1}{\det(\boldsymbol{R})}\begin{bmatrix} r_{2,2} & -r_{1,2} \\ -r_{2,1} & r_{1,1} \end{bmatrix}\begin{bmatrix} a_1 - \frac{1}{2}(r_{11}^2 + r_{12}^2) \\ a_2 - \frac{1}{2}(r_{21}^2 + r_{22}^2) \end{bmatrix} dt \\ &+ \begin{bmatrix} dw_{1,t} \\ dw_{2,t} \end{bmatrix} \end{aligned} \tag{96}$$

and $\tilde{\boldsymbol{Z}}_t \sim N(\boldsymbol{R}^{-1}(\boldsymbol{a} - \frac{1}{2}diag(\boldsymbol{R}\boldsymbol{R}^T))t, \boldsymbol{I}t)$ *and the inverse of* $\tilde{\boldsymbol{Z}}$ *is*

$$\begin{aligned} \boldsymbol{X}_t =& \boldsymbol{\psi}^{-1}(\boldsymbol{Z}_t) = \boldsymbol{\psi}^{-1}(\tilde{\boldsymbol{\psi}}^{-1}(\tilde{\boldsymbol{Z}}_t)) = \boldsymbol{\psi}^{-1}(\boldsymbol{R}\tilde{\boldsymbol{Z}}_t)) = \exp(\boldsymbol{R}\tilde{\boldsymbol{Z}}_t) \\ =& \begin{bmatrix} e^{r_{11}\tilde{Z}_{1,t} + r_{12}\tilde{Z}_{2,t}} \\ e^{r_{21}\tilde{Z}_{1,t} + r_{22}\tilde{Z}_{2,t}} \end{bmatrix}. \end{aligned} \tag{97}$$

□

Theorem 5 states that a process with state independent diffusion can be written as a weighted version of the original process, which has unit diffusion. The weight is the inverse of diffusion matrix, it is tempting to interpret the diffusion matrix ($\boldsymbol{R}(t)$) as a local standard deviation, such an interpretation is however not straight forward, since the local variance is different from $\boldsymbol{R}(t)^2$. The density of an SDE is determined by the Fokker-Planck equation (sometimes referred to as the Kolmogorov forward equation), which in the multidimensional case is given by (Gard, 1988)

$$\begin{aligned} p_t(\boldsymbol{x}, t) =& -\sum_i \frac{\partial}{\partial x_i}\left(f_i(\boldsymbol{x}, t)p(\boldsymbol{x}, t)\right) + \\ & \frac{1}{2}\sum_i\sum_j \frac{\partial^2}{\partial x_i \partial x_j}\sum_k \left(\sigma_{ik}(\boldsymbol{x}, t)\sigma_{jk}(\boldsymbol{x}, t)\right) p(\boldsymbol{x}, t) \end{aligned} \tag{98}$$

$$\begin{aligned} =& -\sum_i \frac{\partial}{\partial x_i}\left(f_i(\boldsymbol{x}, t)p(\boldsymbol{x}, t)\right) + \\ & \frac{1}{2}\sum_i\sum_j \frac{\partial^2}{\partial x_i \partial x_j}\left(\boldsymbol{\sigma}\boldsymbol{\sigma}^T(\boldsymbol{x}, t)\right)_{ij} p(\boldsymbol{x}, t). \end{aligned} \tag{99}$$

The density ($p(\cdot)$) does not depend on $\boldsymbol{\sigma}$ itself, but only on $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$, meaning that p_t is only uniquely determined up to what can loosely be refereed to as the (non unique) "square root" of $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$. However for a positive definite symmetric

matrix, say $\boldsymbol{A}$, there exist a unique positive definite symmetric matrix $\boldsymbol{T}$, such that $\boldsymbol{T}^2 = \boldsymbol{A}$, and by construction $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$ is a positive definite symmetric matrix if $\boldsymbol{\sigma}$ has full rank. Again the best way to understand this is by considering a small example.

Example 10 *Consider the SDE*

$$d\boldsymbol{X}_t = \begin{bmatrix} 2 & -3 \\ -5 & -4 \end{bmatrix} d\boldsymbol{w}_t; \quad \boldsymbol{X}_0 = \boldsymbol{0} \tag{100}$$

it is well known that $\boldsymbol{X}_t$ follow a Gaussian distribution with mean $\boldsymbol{0}$, the variance for this process is given by

$$\begin{aligned} V(\boldsymbol{X}_t) =& \boldsymbol{\sigma}\boldsymbol{\sigma}^T t & (101)\\ =& \begin{bmatrix} 2 & -3 \\ -5 & -4 \end{bmatrix} \begin{bmatrix} 2 & -5 \\ -3 & -4 \end{bmatrix} t & (102)\\ =& \begin{bmatrix} 13 & 2 \\ 2 & 41 \end{bmatrix} t. & (103) \end{aligned}$$

Consider now the proces

$$d\tilde{\boldsymbol{X}}_t = \begin{bmatrix} 3.6 & 0.2 \\ 0.2 & 6.4 \end{bmatrix} d\boldsymbol{w}_t; \quad \boldsymbol{X}_0 = \boldsymbol{0}, \tag{104}$$

the variance of $\tilde{\boldsymbol{X}}_t$ is

$$\begin{aligned} V(\tilde{\boldsymbol{X}_t}) =& \tilde{\boldsymbol{\sigma}}\tilde{\boldsymbol{\sigma}}^T t & (105)\\ =& \begin{bmatrix} 3.6 & 0.2 \\ 0.2 & 6.4 \end{bmatrix}^2 t & (106)\\ =& \begin{bmatrix} 3.6^2 + 0.2^2 & 0.2(6.4 + 3.6) \\ 0.2(6.4 + 3.6) & 6.4^2 + 0.2^2 \end{bmatrix} t & (107)\\ =& \begin{bmatrix} 13 & 2 \\ 2 & 41 \end{bmatrix} t. & (108) \end{aligned}$$

Which shows that $\tilde{\boldsymbol{X}}_t$ is a weak solution to the SDE (100).

Analytic solutions for the unique "square root" of $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$ are not easy to find. This is however not important if we are interested in estimation, since the likelihood is generated by the weak solution to the SDE. The important conclussion is that we can only indetify the number of parameters corresponding to a symetric version of the "square root" of $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$.

In the one dimensional case we can think $\boldsymbol{R}$ as standard deviation is the following sense, let x_t be a continuous time random walk

$$dx_t = r dw_t, \tag{109}$$

then the discrete time stochastic process (with $t_{k-1} < t_k$)

$$y_k = \frac{x_{t_k} - x_{t_{k-1}}}{r\sqrt{t_k - t_{k-1}}}; \quad k = 1, 2, ... \quad , \tag{110}$$

is a sequence of iid. standard Gaussian random variables. The multidimensional equivalent to (109) is

$$d\boldsymbol{x}_t = \boldsymbol{R}d\boldsymbol{w}_t. \tag{111}$$

Now if the matrix $\boldsymbol{R} \in \mathbb{R}^{n\times n}$ is symmetric and positive definite then $\boldsymbol{R}^2 = \boldsymbol{R}\boldsymbol{R}^T$, and the discrete time stochastic process (again with $t_{k-1} < t_k$)

$$\boldsymbol{y}_k = \frac{1}{\sqrt{t_k - t_{k-1}}}\boldsymbol{R}^{-1}(\boldsymbol{x}_{t_k} - \boldsymbol{x}_{t_{k-1}}); \quad k = 1, 2, ... \quad , \tag{112}$$

is a sequence of n-dimensional iid. standard Gaussian random variables, and in this sense we can think of $\boldsymbol{R}$ as the standard deviation. The generation of correlated random variable is often done by simulating independent standard random numbers and then multiplying by the covariation matrix (Madsen, 2008). The transformation in Theorem 5 can be viewed as the SDE equivalent to such a transformation.

The examples presented so far have been rather simple with the purpose to explain or clarify the theory, for such purposes it is not illustrative to include more physical reasoning. It might however be motivating to see an example based on real life reasoning, the last example of this note is such an example.

Example 11 A competition model: *Consider a controlled experiment with two living populations P_1 and P_2 (e.g. bacteria or phytoplankton) eating the same two nutrients N_1 and N_2 (e.g. nitrogen and phosphor), but not each other. Let the experiment be constructed such the the total amount of nutrients are held constant. Biological growth models are often assumed to follow Liebigs law of minimum and Michaelis-Menten kinetics, i.e.*

$$dP_i = \left(\min\left(\frac{\mu_{i,1}N_1}{k_{i,1} + N_1}, \frac{\mu_{i,2}N_2}{k_{i,2} + N_2}\right) - m_i\right) P_i dt \tag{113}$$

$$= (f_i(\boldsymbol{N}) - m_i)P_i dt, \tag{114}$$

where $m_i > 0$ is the mortality rate and $\min(\cdot)$ *express the limiting factor. Further let a_{ij} be conversion factors that convert population i to nutrient j, such factors are often known or approximately known from literature. As discussed earlier the diffusion term for biological processes is often assumed to be proportional to either P_i of $\sqrt{P_i}$, here we assume that the diffusion is proportional to $P_i^{\gamma_i}$ with $\gamma_i \in (\frac{1}{2}, 1)$ and leave it to the estimation procedure to determine γ_i.*

The SDE for the system described above is

$$d\begin{bmatrix} N_{1,t} \\ N_{2,t} \\ P_{1,t} \\ P_{2,t} \end{bmatrix} = \begin{bmatrix} -a_{11}(f_1(\boldsymbol{N}) - m_1) & -a_{12}(f_2(\boldsymbol{N}) - m_2) \\ -a_{21}(f_1(\boldsymbol{N}) - m_1) & -a_{22}(f_2(\boldsymbol{N}) - m_2) \\ f_1(\boldsymbol{N}) - m_1 & 0 \\ 0 & f_2(\boldsymbol{N}) - m_2 \end{bmatrix} \begin{bmatrix} P_{1,t} \\ P_{2,t} \end{bmatrix} dt + \begin{bmatrix} -a_{11}\sigma_1 P_{1,t}^{\gamma_1} & -a_{12}\sigma_2 P_{2,t}^{\gamma_2} \\ -a_{21}\sigma_1 P_{1,t}^{\gamma_1} & -a_{22}\sigma_2 P_{2,t}^{\gamma_2} \\ \sigma_1 P_{1,t}^{\gamma_1} & 0 \\ 0 & \sigma_2 P_{2,t}^{\gamma_2} \end{bmatrix} \begin{bmatrix} dw_{1,t} \\ dw_{2,t} \end{bmatrix}. \tag{115}$$

Seemingly we cannot apply the derived methods to transform this system to a system with constant diffusion, however the above system have a 2-dimensional distribution only and transformation is therefore possible. Using Theorem 4 on P_i *gives*

$$\tilde{P}_{i,t} = \psi_i(P_{i,t}) = \frac{1}{\sigma_i} \int x_i^{-\gamma_i} dx \Big|_{x=P_{i,t}} = \frac{P_{i,t}^{1-\gamma_i}}{\sigma_i(1-\gamma_i)}, \tag{116}$$

with the inverse function given by

$$P_{i,t} = (1-\gamma_i)^{\gamma_i - 1}(\sigma_i \tilde{P}_{i,t})^{\gamma_i - 1} = (\tilde{\gamma}_i \sigma_i \tilde{P}_{i,t})^{-\tilde{\gamma}_i}, \tag{117}$$

where $\tilde{\gamma}_i = 1 - \gamma_i$*. Now choose*

$$\tilde{N}_{i,t} = \phi_i(N_{i,t}, P_{1,t}, P_{2,t}) = N_{i,t} + a_{i1}P_{1,t} + a_{i2}P_{1,t}. \tag{118}$$

Using Itô's formula we get

$$\begin{aligned} d\tilde{N}_{i,t} =& (-a_{i1}(f_1(\boldsymbol{N}) - m_1)P_{1,t} - a_{i2}(f_2(\boldsymbol{N}) - m_2)P_{2,t})dt \\ & - a_{i1}\sigma_1 P_{1,t} dw_{1,t} - a_{i2}\sigma_2 P_{2,t} dw_{2,t} \\ & + a_{i1}(f_1(\boldsymbol{N}) - m_1)P_{1,t}dt + a_{i1}\sigma_1 P_{1,t} dw_{1,t} \\ & + a_{i2}(f_2(\boldsymbol{N}) - m_2)P_{2,t}dt + a_{i2}\sigma_2 P_{2,t} dw_{2,t} && (119) \\ =& 0. && (120) \end{aligned}$$

Implying that $\tilde{N}_{i,t} = \tilde{N}_{i,0}$ *is constant and* $N_{i,t}$ *is given by*

$$\begin{aligned} N_{i,t} =& N_i(\boldsymbol{P}_t) && (121) \\ =& \tilde{N}_{i,t} - a_{i1}P_{1,t} - a_{i2}P_{2,t} && (122) \\ =& \tilde{N}_{i,0} - a_{i1}(\tilde{\gamma}_1\sigma_1\tilde{P}_{1,t})^{-\tilde{\gamma}_1} - a_{i2}(\tilde{\gamma}_2\sigma_2\tilde{P}_{2,t})^{-\tilde{\gamma}_2}. && (123) \end{aligned}$$

The system equation for the transformed system takes the form

$$d\begin{bmatrix} \tilde{P}_{1,t} \\ \tilde{P}_{2,t} \end{bmatrix} = \begin{bmatrix} (\sigma_1\tilde{\gamma}_1\tilde{P}_{1,t})^{-\tilde{\gamma}_1}(\tilde{f}_1(\tilde{\boldsymbol{P}}) - m_1) \\ (\sigma_2\tilde{\gamma}_2\tilde{P}_{2,t})^{-\tilde{\gamma}_2}(\tilde{f}_2(\tilde{\boldsymbol{P}}) - m_2) \end{bmatrix} dt + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} dw_{1,t} \\ dw_{2,t} \end{bmatrix}, \tag{124}$$

with

$$\tilde{f}_i(\tilde{\boldsymbol{P}}) = \min\left(\frac{\mu_{i,1}N_1(\tilde{\boldsymbol{P}})}{k_{i,1}+N_1(\tilde{\boldsymbol{P}})}, \frac{\mu_{i,2}N_2(\tilde{\boldsymbol{P}})}{k_{i,2}+N_2(\tilde{\boldsymbol{P}})}\right) \tag{125}$$

$$= \min\left(\frac{\mu_{i,1}(\tilde{N}_{1,0} - a_{11}(\tilde{\gamma}_1\sigma_1\tilde{P}_{1,t})^{-\tilde{\gamma}_1} - a_{12}(\tilde{\gamma}_2\sigma_2\tilde{P}_{2,t})^{-\tilde{\gamma}_2})}{k_{i,1}+\tilde{N}_{1,0} - a_{11}(\tilde{\gamma}_1\sigma_1\tilde{P}_{1,t})^{-\tilde{\gamma}_1} - a_{12}(\tilde{\gamma}_2\sigma_2\tilde{P}_{2,t})^{-\tilde{\gamma}_2}},\right.$$
$$\left.\frac{\mu_{i,2}(\tilde{N}_{2,0} - a_{21}(\tilde{\gamma}_1\sigma_1\tilde{P}_{1,t})^{-\tilde{\gamma}_1} - a_{22}(\tilde{\gamma}_2\sigma_2\tilde{P}_{2,t})^{-\tilde{\gamma}_2})}{k_{i,2}+\tilde{N}_{2,0} - a_{21}(\tilde{\gamma}_1\sigma_1\tilde{P}_{1,t})^{-\tilde{\gamma}_1} - a_{22}(\tilde{\gamma}_2\sigma_2\tilde{P}_{2,t})^{-\tilde{\gamma}_2}}\right). \tag{126}$$

The derivations above strongly depend on the fact that the actual dimension of the joint distribution at time t is only 2, if there had been a random input of nutrient to the system, the derivation would not have been possible. It is therefore a crucial assumption that the experiment is conducted in a controlled environment, with no random interactions with the surroundings.

For the sake of completeness we will give an example of the observation equation, where we will assume that we are able to observe all the states of the original system, and that these observations are log*-normally distributed around the true state, i.e.*

$$\begin{bmatrix} Y_{N_1,k} \\ Y_{N_2,k} \\ Y_{P_1,k} \\ Y_{P_2,t_k} \end{bmatrix} = \begin{bmatrix} \log(N_{1,t_k}) \\ \log(N_{2,t_k}) \\ \log(P_{1,t_k}) \\ \log(P_{2,t_k}) \end{bmatrix} + \begin{bmatrix} \epsilon_{N_1,t} \\ \epsilon_{N_2,t} \\ \epsilon_{P_1,t} \\ \epsilon_{P_2,t} \end{bmatrix} \tag{127}$$

$$= \begin{bmatrix} \log(\tilde{N}_{1,0} - a_{11}(\tilde{\gamma}_1\sigma_1\tilde{P}_{1,t_k})^{-\tilde{\gamma}_1} - a_{12}(\tilde{\gamma}_2\sigma_2\tilde{P}_{2,t_k})^{-\tilde{\gamma}_2}) \\ \log(\tilde{N}_{2,0} - a_{21}(\tilde{\gamma}_1\sigma_1\tilde{P}_{1,t_k})^{-\tilde{\gamma}_1} - a_{22}(\tilde{\gamma}_2\sigma_2\tilde{P}_{2,t_k})^{-\tilde{\gamma}_2}) \\ -\tilde{\gamma}_1(\log(\tilde{\gamma}_1) + \log(\sigma_1) + \log(\tilde{P}_{1,t_k})) \\ -\tilde{\gamma}_2(\log(\tilde{\gamma}_2) + \log(\sigma_2) + \log(\tilde{P}_{2,t_k})) \end{bmatrix} +$$
$$\boldsymbol{\epsilon}_t, \tag{128}$$

where $\boldsymbol{\epsilon}_{t_k}$ *follow a Gaussian distribution with mean zero variance* $\boldsymbol{S}$*.* □

5 Summary and conclusion

We have shown how a class SDE's with state dependent diffusion can be transformed into SDE's with state independent diffusion. For one dimensional systems this transformation is rather straight forward and is only limited by the ability to find a closed form inverse transformation. Such transformations are

important both in estimation and simulations. Iacus (2008) notes that the Lamperti transformation or similar transformations (not necessarily to unity) should always be used before simulation and that many estimation techniques rely on unit or constant diffusion. Luschgy (2006) presents proofs of convergence rates for a simulations procedure, which also relies on the existence of the Lamperti transform.

For time dependent diffusion the transformed process will depend on the time derivative of the transformation, which is equivalent to dependence on the time-derivative of the diffusion term. While this might be reasonable when the functional relation between the diffusion and time is given in an explicit form, it is problematic if the time dependence on the diffusion is through an observed input, because numerical differentiation will be needed.

For multidimensional diffusion processes the transformation to systems with state independent diffusion is more delicate, and Luschgy (2006) note that the Lamperti transform is essentially a one-dimensional transformation. This is also what has been shown here, however, it is also stressed that even with the restriction given in Theorem 4, there is still a large class of SDE's that can be handled through transformations. This class includes processes that seemingly is not included in Theorem 4, like the mass balance model presented in Example 11.

It is shown that the transformation to unit diffusion can be interpreted as a weighting with the local standard deviation. This means that the system noise innovation is equal for all states in the transformed process.

References

Aït-Sahalia Y. (2008) Closed-form likelihood expansions for multivariate diffusions. *The Annals of Statistics*, **32**, (2), 906-937

Forman, J., and Sørensen, M. (2008) The Pearson Diffusion: A Class of Statistically Tractable Diffusion Processes. *Scandinavian Journal of Statics*, **35**, 438-465

Gard T.C. (1988) Introduction to stochastic differential equations. *Marcel Dekker, Inc.*, New York.

Iacus S. M. (2008) Simulation and Inference for Stochastic Differential Equations - With R Examples. *Springer Series in Statistics*

Jazwinski A.H. (1970) Stochastic Processes and Filtering Theory. *Academic Press*, New York.

Klebaner F. C. (2005) Introduction to Stochastic Calculus with Applications., *Imperial College Press*

Kristensen N.R., and Madsen H. (2003) Continuous time stochastic modeling - CTSM 2.3 - Mathematics Guide. *Technical University of Denmark*

Kristensen N.R., Madsen H., and Jørgensen S.B. (2004) Parameter estimation in stochastic grey-box models. *Automatica*, **40**, 225-237

Luschgy H., and Pagés G. (2006) Functional quantization of a class of Brownian diffusions: A constructive approach. *Stochastic Processes and their Applications* **116***, 310-336*

Madsen, H. (2008) *Time Series Analysis.* Chapman & Hall/CRC

Nielsen J.N., and Madsen H. (2001) Applying the EKF to stochastic differential equations with level effects. *Automatica,* **37***, 107-112*

Øksendal B. (2003) Stochastic Differential Equations - An Introduction with Applications, Sixth edition. *Springer, Berlin*

Paper E

Applying EKF for Maximum Likelihood Estimation in Stochastic Differential Equations with State Dependent Diffusion

---

Authors:
J.K. Møller, and H. Madsen.

To be submitted to:
Automatica

Remarks:
The paper was not submitted when the thesis was handed in and minor changes is expected before the final journal submision.

Applying EKF for Maximum Likelihood Estimation in Stochastic Differential Equations with State Dependent Diffusion

Jan Kloppenborg Møller[1,2], and Henrik Madsen[1]

Abstact

Transformations of Stochastic Differential Equations (SDEs) that leaves the diffusion term independent of the state are often referred to as Lamperti transformations, while the Lamperti transform always exist for univariate SDEs, the existence in the multivariate case depend on the structure of the diffusion matrix. In the present study we correct an earlier reported result (Nielsen & Madsen, 2001) for multivariate SDEs. The obtained formulation of a multivariate Lamperti transformation for a restricted class of SDEs, is used to demonstrate that reliable parameter estimates can be obtained by likelihood estimation based on the Extended Kalman Filter (EKF). The divergence between direct application of EKF and exact Bayesian filtering is demonstrated for the CIR-model. The same analysis show the good results can be expected when using the EKF combined with the Lamperti transform. More complex simulation examples of two-dimensional Lotka-Volterra systems show that reliable parameter estimates can still be expected when the system is subject to strong random forcing introduced by the diffusion term.

Key Words: Stochastic differential equations, State dependent diffusion, Lamperti transform, Extended Kalman filter, Maximum likelihood estimation, Stochastic Lokta-Volterra systems.

1 Introduction

The Kalman filter has enjoyed enormous success since the introduction in 1960 (Kalman, 1960), the development to the Extended Kalman Filter (EKF) (Jazwinski, 1970) enabled approximate likelihood estimation for non-linear stochastic

[1]DTU Informatics, Richard Pedersens Plads, Technical University of Denmark Building 321, DK-2800 Lyngby, Denmark.

[2]National Environmental Research Institute, Fredriksborgvej 399, DK-4000 Roskilde, Denmark.

systems. The present study is restricted stochastic differential equations (SDEs) and in this case the stochastic perturbations are differential Wiener noise multiplied by a diffusion coefficient. The Extended Kalman filter applies as a second order moment approximation for general structures of the stochastic perturbations (Simon, 2006). Gaussian assumptions on the transition probabilities will, however, depend on correlation structures and many likelihood methods depend on transformations of the state space which leaves the diffusion term independent of the state (Iacus, 2008). This transformation is often referred to as the Lamperti transform (Iacus, 2008; Luschgy & Pagés, 2006). For one-dimensional SDEs the Lamperti-transform always exist, while for multi-dimensional SDEs it exists for a restricted class of SDEs only, Aït-Sahalia (2008) refer to this class as reducible SDEs.

The EKF presented by Jazwinski (1970) and the implementation used in the simulation examples of the present study assume state independent diffusion (Kristensen et al., 2004). Even though the Lamperti transform can only be applied to limited the class of multivariate SDE models, the remaining class of models is still very rich compared to the class of SDE-model with state independent diffusion. Though more sophisticated methods, like particle filters (e.g. Givon et al., 2009) exists, EKF based methods are still relevant, as it does not depend on simulation and therefore enables faster evaluations, a further advantages is that easily accessible software exist (e.g. CTSM[1], Kristensen & Madsen (2003); Kristensen et al. (2004)).

Although the structural demands on the diffusion term are well-known (Aït-Sahalia, 2008; Luschgy & Pagés, 2006), some confusion still remains in more applied settings. Nielsen & Madsen (2001) generalise the one dimensional result reported by Baadsgaard et al. (1997), the application of Itô's Lemma is however wrong and the present study start out by summarising the results in Nielsen & Madsen (2001) followed by the corrections needed to give a correct formulation. The application of the resulting formulation is illustrated by simulation studies of two-dimensional stochastic Lotka-Volterra systems, the analysis include a setup, which is directly within the presented class of models and further an example which does not belong to the presented class of models, but where the reducibly is obtained, because the dimension of the stochastic perturbations is one. The examples also illustrate how the transformation can be used to analyse how diffusion affect the dynamics of the system.

[1] www.imm.dtu.dk/ ctsm

2 Comments on an earlier reported result

We start by an analysis of the argumentation given in Nielsen & Madsen (2001). This section goes through the argumentation and presents the corrections needed to get a correct formulation.

Given the system of stochastic differential equations

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t; \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{x}_t, t; \boldsymbol{\theta})d\boldsymbol{w}_t, \tag{1}$$

where $\boldsymbol{x}_t \in \mathbb{R}^n$ is the state, $\boldsymbol{u}_t \in \mathbb{R}^l$ is the input, and $\boldsymbol{\theta} \in \mathbb{R}^p$ is a parameter vector to be estimated. The task is to find a transformation

$$\boldsymbol{z}_t = \boldsymbol{\psi}(\boldsymbol{x}_t, t; \boldsymbol{\theta}), \tag{2}$$

such that

$$d\boldsymbol{z}_t = \tilde{\boldsymbol{f}}(\boldsymbol{z}_t, \boldsymbol{u}_t, t; \boldsymbol{\theta})dt + \tilde{\boldsymbol{\sigma}}(t; \boldsymbol{\theta})d\boldsymbol{w}_t, \tag{3}$$

where $\tilde{\boldsymbol{\sigma}}$ is independent of the state $\boldsymbol{z}_t$.

The claim in Nielsen & Madsen (2001) is that with some additional assumptions this can be achieved. The assumptions are:

Assumption 1 *Assume that the diffusion term is restricted to nonzero, i.e.*

$$\sigma^{ij}(\boldsymbol{x}_t, t; \boldsymbol{\theta}) \neq 0, \quad i = 1, ..., n, j = 1, ..., m. \tag{4}$$

Assumption 2 *Assume that for each i there exist only one σ^{ij} as a function of one and only one state variable $\boldsymbol{x}_t^{\nu(i)}$, where $\nu(i)$ should be different for each i, i.e.*

$$\sigma^{ij}(\boldsymbol{x}_t, t; \boldsymbol{\theta}) = \sigma^{ij}(x_t^{\nu(i)}, t; \boldsymbol{\theta}), \quad i = 1, ..., n, j = 1, ..., m. \tag{5}$$

Assume further that $\sigma^{ij}(x_t^{\nu(i)}, t; \boldsymbol{\theta})$ is bijective and that the function $[\sigma^{ij}(x, t; \boldsymbol{\theta})]^{-1}$ is integrable with respect to x.

Then, according to Nielsen & Madsen (2001), the following theorem apply

Theorem 1 *Let $\boldsymbol{x}_t$ be a solution to (1). Then Assumptions 1 and 2 provide necessary and sufficient conditions for the existence of a transformation (2) given by*

$$\psi^k(x_t^{\nu(i)}, t) = \int \frac{d\xi}{\sigma^{ij}(\xi, t)}\bigg|_{\xi = x_t^{\nu(i)}}, \quad k, i = 1, ...n, j = 1, ..., m \tag{6}$$

such that (3) is fulfilled.

Assumption 2 imply that

$$\sigma^{i1}(\boldsymbol{x}_t,t,\boldsymbol{\theta})=\sigma^{i2}(x_t^{\nu(i)},t,\boldsymbol{\theta})=...=\sigma^{im}(x_t^{\nu(i)},t,\boldsymbol{\theta})=\sigma^{i}(x_t^{\nu(i)},t,\boldsymbol{\theta}),\quad i=1,...,n, \tag{7}$$

and further the $i=k$ in Theorem 1 and the transformation (6) reads

$$\psi^k(x_t^{\nu(k)},t)=\int\frac{d\xi}{\sigma^{kj}(\xi,t)}\bigg|_{\xi=x_t^{\nu(k)}},\quad i=1,...n,j=1,...,m. \tag{8}$$

The proof of Theorem 1 was completed by writing Itô's multivariate Lemma and claiming that the result follows immediately when $\nu(i)\neq i$. A closer inspection do however show that the conditions are fulfilled only if $\nu(i)=i$. By Itô's Lemma we get

$$\begin{aligned}dz_t^k=&\frac{\partial\psi^k}{\partial t}(\boldsymbol{x}_t,t)dt+\sum_{i=1}^n\frac{\partial\psi^k}{\partial x_i}(\boldsymbol{x}_t,t)dx_t^i\\&+\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n\frac{\partial^2\psi^k}{\partial x_i\partial x_j}(\boldsymbol{x}_t,t)dx_t^jdx_t^i \qquad (9)\\=&\left(\psi_t^k+\sum_{i=1}^n\psi_{x^i}^kf^i\right)dt+\sum_{i=1}^n\psi_{x_i}^k\left(\sum_{h=1}^m\sigma^{ih}dw_t^h\right)+\\&\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n\psi_{x_i,x_j}^k\left(\sum_{h=1}^m\sigma^{j,h}dw_t^h\right)\left(\sum_{l=1}^m\sigma^{i,l}dw_t^l\right), \qquad (10)\end{aligned}$$

since $\psi_{x^i}^k=0$ for $i\neq\nu(k)$, Eq. (10) simplifies to

$$\begin{aligned}dz_t^k=&\left(\psi_t^k+\psi_{x^{\nu(k)}}^kf^{\nu(k)}\right)dt+\psi_{x^{\nu(k)}}^k\sum_{h=1}^m\sigma^{\nu(k),h}dw_t^h\\&+\frac{1}{2}\psi_{x^{\nu(k)},x^{\nu(k)}}^k\sum_{h=1}^m\left(\sigma^{\nu(k),h}\right)^2dt \qquad (11)\\=&\tilde{f}_kdt+\psi_{x^{\nu(k)}}^k\sum_{h=1}^m\sigma^{\nu(k),h}dw_t^h, \qquad (12)\end{aligned}$$

inserting (8) gives

$$\begin{aligned}dz_t^k=&\tilde{f}_kdt+\frac{\partial}{\partial x^{\nu(k)}}\int\frac{d\xi}{\sigma^k(\xi,t)}\bigg|_{\xi=x_t^{\nu(k)}}\sum_{h=1}^m\sigma^{\nu(k)}dw_t^h \qquad (13)\\=&\tilde{f}_kdt+\frac{1}{\sigma^k(x_t^{\nu(k)},t)}\sigma^{\nu(k)}\left(x_t^{\nu^2(k)},t\right)\sum_{h=1}^m dw_t^h, \qquad (14)\end{aligned}$$

if σ^k is not a constant the diffusion of (14) can only be independent of the state if $\nu^2(k) = \nu(k) \Rightarrow \nu(k) = k$, and this completes the proof.

Assumption 1 is also not a necessary condition, and it excludes important special cases, such as

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t)dt + \text{diag}(\boldsymbol{x}_t)\text{diag}(\boldsymbol{\sigma})d\boldsymbol{w}_t, \tag{15}$$

where $\boldsymbol{\sigma}$ is a constant vector.

It follows from the analysis that we can state the following general result

Theorem 2 *Let $\boldsymbol{x}_t$ be an Itô SDE given by*

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, t)dt + \boldsymbol{\sigma}(\boldsymbol{x}_t, t)\boldsymbol{R}(t)d\boldsymbol{w}_t, \tag{16}$$

where $\boldsymbol{R}(t) \in \mathbb{R}^{n\times n}$ is any matrix function of t, and $\boldsymbol{\sigma}(\boldsymbol{x}_t, t) \in \mathbb{R}^{n\times n}$ is a diagonal matrix with diagonal elements $\sigma^{i,i}(\boldsymbol{x}_t, t)$ given by

$$\sigma^{i,i}(\boldsymbol{x}_t, t) = \sigma^i(x_t^i, t). \tag{17}$$

Assume further that $\sigma^i(x_t^i, t; \boldsymbol{\theta})$ is bijective and that the function $[\sigma^i(x, t; \boldsymbol{\theta})]^{-1}$ is integrable with respect to x. Then the transformations

$$z_t^i = \psi^i(x_t^i, t) = \int \frac{d\xi}{\sigma^i(\xi, t)}\bigg|_{\xi=x_t^i}, \tag{18}$$

will result in an Itô SDE with the i'th element given by

$$\begin{aligned} dz_t^i = \Bigg(&\frac{\partial}{\partial t}\psi^i(x, t)\big|_{x=\psi_i^{-1}(z_t^i,t)} + \frac{f^i(\boldsymbol{\psi}^{-1}(\boldsymbol{z}_t, t), t)}{\sigma^i(\psi_i^{-1}(z_t^i, t), t)} \\ &-\frac{1}{2}\sigma_x^i(\psi_i^{-1}(z_t^i, t), t)\Bigg) dt + \sum_{j=1}^{n} r^{i,j}(t)dw_t^j. \end{aligned} \tag{19}$$

with $r^{i,j}(t)$ elements of $\boldsymbol{R}(t)$, ψ_i^{-1} the inverse of ψ^i, and

$$\boldsymbol{x}_t = \boldsymbol{\psi}^{-1}(\boldsymbol{z}_t, t). \tag{20}$$

The proof of the theorem follows the methodology of the analysis above, and will not be presented here, but is available in Møller & Madsen (2010). The results is a time dependent generalisation of the result reported in Luschgy & Pagés (2006), and as noted in the same reference the application of Itô's Lemma is essentially one-dimensional. It should also be noted that there exists more general results than Theorem 2 (Aït-Sahalia, 2008), we will however use the

result stated above to illustrate that reliable parameter estimation is available through Maximum Likelihood estimation and EKF, at least for moderate values of the diffusion term. Further, as we will show in Section 3 and 4, the drift term of the Lamperti transformed process can help to identify upper bounds for the parameters that can be identified through ML-estimation by the EKF.

3 EKF and Maximum Likelihood estimation for processes with state dependent diffusion

When applying the EKF for likelihood estimation the explicit assumption is that the distributions of the one-step state prediction and the state reconstruction are both well described by the second order moment representation and that the mode and the expectation are close. A consequence of Dynkins formula (cf. Øksendal (2003)) and the martingale property of Itô-integrals is that the expectation of a stochastic process generated by an Itô-SDE is independent of the diffusion term. That this might be a very poor estimate of the mode can be realised from simple SDEs, like stochastic population growth models with the diffusion linear in the state (Øksendal, 2003)

$$dx_t = ax_t dt + \sigma x_t dw_t, \tag{21}$$

at time t the solution to this model is a log-normal distribution with the mean and variance both depending on time and $\frac{E[x_t]}{\text{mode}(x_t)} = e^{\frac{3}{2}\sigma^2 t}$. Implying that the divergence between the expectation and the mode is exponential in t.

The implementation of Maximum Likelihood estimation based on EKF we use here is described in Kristensen et al. (2004) and the details will not be presented here. However, the central assumption is that the system is described by the continuous-discrete time state-space formulation

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, t; \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{u}_t, t; \boldsymbol{\theta})d\boldsymbol{w}_t \tag{22}$$

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_{t_k}, \boldsymbol{u}_{t_k}, t_k; \boldsymbol{\theta}) + \boldsymbol{e}_k, \tag{23}$$

where $\boldsymbol{e}_k$ is a Gaussian random variable with expectation zero and variance $\boldsymbol{S}$. Further the assumption is that the one-step state prediction is Gaussian.

The implementation requires the diffusion term to be independent of the state, and for computational convenience alone it might be an advantage to apply the presented transformation. However, as illustrated in the example below it might also give much better state estimation, in particular when combined with a non-linear observation equation. Non-linear observations might produce state

estimates which are outside the natural state space for the stochastic process. If the natural state space is assumed in the observation equation (e.g. the log-transform of a growth model), then evaluation of the likelihood might be impossible. The example presented below illustrates these points.

3.1 Example 1: State estimation - A simple example

Consider the CIR-model (Iacus, 2008)

$$dx_t = (\theta_1 - \theta_2 x_t)dt + \theta_3\sqrt{x_t}dw_t, \quad x_0 = \frac{1}{2}, \tag{24}$$

with $\theta_1 = \theta_2 = 1$. The behaviour will be analysed with different values of θ_3. We assume that the process is observed in discrete time and that the observations are log-normal distributed around the true state

$$\log(y_k) = \log(x_{t_k}) + e_k, \tag{25}$$

where $e_k \sim N(0, 0.1)$, and at time $t = 1$ the observed value is $y_1 = \frac{1}{4}$.

Below we compare state estimation for the CIR model by direct use of the EKF, EKF applied to the Lamperti transformed process, and exact Bayesian filtering for different values of θ_3.

Exact Bayesian filtering:

The CIR model is simple enough to have a known transition probability distribution, which can be calculated from a non-central χ^2-distribution (Iacus, 2008). It is therefore possible to compare estimation by the Bayesian recursive filter and direct estimation via the EKF and estimation via the Lamperti transformation and the EKF. The Bayesian filter cannot be evaluated when $\theta_3 > \sqrt{2\theta_1}$, because the conditional density has mode equal zero with $p(x_t = 0|t, x_0) = \infty$.

Direct state estimation by EKF:

The conditional mean and variance of (24) with the given parameters are (Iacus, 2008)

$$\hat{x}_{t|0} = 1 + (x_0 - 1)e^{-t} \tag{26}$$

$$P_{t|0} = \theta_3^2 \left(x_0 e^{-t}(1 - e^{-t}) + \frac{1}{2}(1 - e^{-2t}) \right), \tag{27}$$

in implementations of the EKF these values might be obtained by approximate method like sub-sampling or the iterative EKF (Kristensen et al., 2004). Eqs. (26) - (27) are used to calculate the reconstruction for different values of θ_3 (Figure 1).

State estimation by EKF and the Lamperti transform:

The Lamperti-transform for the process (24) is

$$z_t = \psi(x_t) = 2\sqrt{x_t} \quad \Rightarrow x_t = \psi^{-1}(z_t) = \frac{1}{4}z_t^2, \tag{28}$$

and z_t is governed by the Itô-SDE

$$dz_t = \left[\frac{2}{z_t}\left(\theta_1 - \frac{1}{4}\theta_3^2\right) - \frac{1}{2}\theta_2 z_t\right]dt + \theta_3 dw, \tag{29}$$

to have a meaningful interpretation of (29) in the limit $z_t \to 0$, we need $\theta_3^2 < 4$. The conditional mean and variance of z_t is given as solutions to the ODEs (Kristensen et al., 2004)

$$\frac{d\hat{z}_{t|0}}{dt} = \frac{2}{\hat{z}_{t|0}}\left(\theta_1 - \frac{1}{4}\theta_3^2\right) - \frac{1}{2}\theta_2\hat{z}_{t|0} + \theta_3^2 \tag{30}$$

$$\frac{P^z_{t|0}}{dt} = \left[\frac{\partial}{\partial z}\left(\frac{2}{z}\left(\theta_1 - \frac{1}{4}\theta_3^2\right) - \frac{1}{2}\theta_2 z\right)\bigg|_{z=\hat{z}_{t|0}}\right]^2 P^z_{t|0} + \theta_3^2 \tag{31}$$

$$= \left[-\frac{2}{\hat{z}^2_{t|0}}\left(\theta_1 - \frac{1}{4}\theta_3^2\right) - \frac{1}{2}\theta_2\right]^2 P^z_{t|0} + \theta_3^2, \tag{32}$$

with initial conditions

$$z_0 = 2\sqrt{x_0}, \quad P^z_{t|0} = 0. \tag{33}$$

With the initial conditions (33) and the chosen parameters the solution to (30) and (32) can be found as

$$\hat{z}_{t|0} = 2\sqrt{x_0 e^{-t} + \left(1 - \frac{\theta_3^2}{4}\right)(1 - e^{-t})} \tag{34}$$

$$P^z_{t|0} = \frac{\theta_3^2}{2}\frac{2x_0 e^{-t}(1 - e^{-t}) + \left(1 - \frac{\theta_3^2}{4}\right)(1 - e^{-t})^2}{x_0 e^{-t} + \left(1 - \frac{\theta_3^2}{4}\right)(1 - e^{-t})}, \tag{35}$$

this second order moment representation is used together with the observation equation

$$y_k = \log(x_{t_k}) + e_k \tag{36}$$
$$= 2\log(z_{t_k}) - \log(4) + e_k, \tag{37}$$

to obtain the reconstruction and the prediction.

Comparing the results:

The state estimation from the three different filter approximations are given in Figure 1. When θ_3 is large, the direct application of the EKF gives poor state estimation results with very large variance. Further the estimation produce negative state estimation, which imply that the filter prediction is does not belong to the state space of x_t. If the next observation is obtain at time t_2 is close to t_1, then the observation equation cannot be evaluated.

Applying the EKF to the Lamperti transformed process gives more convincing results and the Bayesian filter reconstruction and the EKF reconstruction are close for the entire range of values where the Bayesian filter reconstruction can be evaluated. Even though the mode is quite closely captured by the EKF, the confidence bands are not well captured by the Lamperti EKF and the Gaussian assumption on the transformed process is therefore more doubtful. Still the result is much better than direct application of the EKF.

Further the one-step prediction, which is needed to evaluate the likelihood is independent of θ_3 for the direct application of the EKF, while the one-step prediction by the Bayesian filter dependent strongly on the value of θ_3. The Lamperti EKF state prediction depend on θ_3, and even though the prediction will be off for large values of θ_3 it is much closer to the exact Bayesian filter prediction.

The example illustrates that the Lamperti transform should not only be applied for computational convenience, but also to get more correct state estimates by the EKF. The following section present simulation based examples of stochastic Lokta-Voltarre models and illustrates that maximum likelihood estimation based on EKF state estimates can successfully be applied to multivariate systems also when the diffusion term dominates the dynamics.

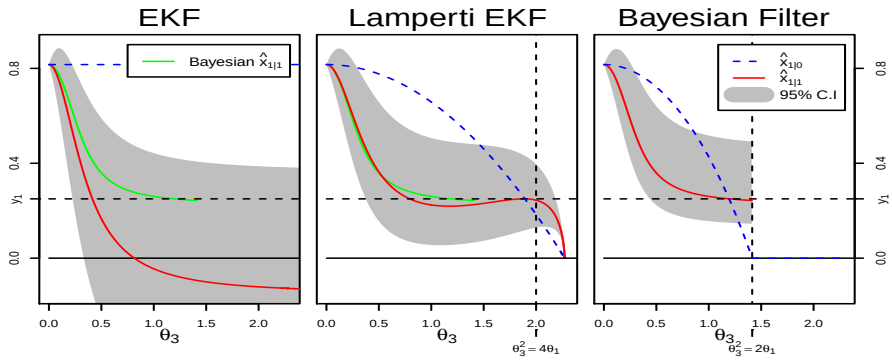

Figure 1: State predictions and state reconstructions of the SDE (24) and the observation equation (25), with $x_0 = \frac{1}{2}$. "EKF" refer to direct application of the EKF, "Lamperti EKF" refer to application of the EKF to the Lamperti-transform process (29) with the observation equation (37) and "Bayesian Filter" refer to the exact recursive Bayesian Filter estimates.

4 Simulation examples - Stochastic Lokta-Voltarre type models

The stochastic Lotka-Voltarre model makes an interesting case study, because it is one of the simplest non-linear multidimensional systems we can think of. The simplicity implies that we can compare the dynamics of the deterministic and the stochastic model. The comparison illustrate that the inclusion of stochastic perturbation have a profound influence on the long-term dynamics of the system. A SDE generalisation of a simple two-dimensional Lotka-Voltarre system is given by (see Klebaner (2005) for an example with Poisson noise)

$$d\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} ax_{1,t} - bx_{1,t}x_{2,t} \\ cx_{1,t}x_{2,t} - dx_{2,t} \end{bmatrix} dt + \boldsymbol{\sigma}(\boldsymbol{x}_t)d\boldsymbol{w}, \tag{38}$$

in a biological context we can think of $x_{1,t}$ as the abundance of herbivores and $x_{2,t}$ and the abundance of carnivores, a is the birth-rate of herbivores, d is the death-rate of carnivores, b is the predation and $\frac{c}{b}$ is the number of herbivores needed to produce one carnivore, the stationary point of the deterministic skeleton is $\boldsymbol{x}_0 = \left[\frac{d}{c}, \frac{a}{b}\right]^T$. The idea of this section is to illustrate maximum likelihood estimation in nonlinear SDE models with state dependent noise, and although we give some remarks on the noise formulation, we do not attempt to give realistic ecosystem interpretations of the parameters, for such analysis see e.g. Gilioli et al. (2008).

To illustrate the estimation methodology we need to give the continuous-discrete time state-space formulation, which links a set of observation to the system equation (38), in the presented simulation examples we will use the observation equation

$$\boldsymbol{y}_k = \log(\boldsymbol{x}_{t_k}) + \boldsymbol{e}_k, \tag{39}$$

where $\boldsymbol{y}_k$ is the observation at time t_k, and $\boldsymbol{e}_k$ is a Gaussian random variable with expectation zero and covariance $\boldsymbol{S}$. In all examples below $\boldsymbol{S}$ is given by

$$\boldsymbol{S} = \begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix}, \tag{40}$$

and the system is simulated for t in $[0, 10]$, and observations are taken at times $t_k = \{0, 0.1, .., 10\}$.

The structure of the diffusion term should be decided based on mathematical and biological considerations of the system. A basic requirement on this system is that $P(\boldsymbol{x_t} < 0) = 0 \; \forall \; t$, which excludes pure additive diffusion. Clearly diffusions that fulfil this requirement are numerous, here we will treat two simple cases; one with stochastic death and birth rate and one with stochastic predation.

Case 1: Stochastic birth and death rates: If only birth and death rates are stochastic then a simple diffusion is

$$\boldsymbol{\sigma}(\boldsymbol{x}_t) = \begin{bmatrix} \sigma_1 x_{1,t}^{\gamma_1} & 0 \\ 0 & \sigma_2 x_{2,t}^{\gamma_2} \end{bmatrix} \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}, \tag{41}$$

where γ_i controls how the diffusion scales with the states and r determines the noise correlation. The properties of the model for different choices of γ_i will be discussed below.

Case 2: Stochastic predation: If only predation is stochastic then a natural diffusion model is (Gilioli et al., 2008)

$$\boldsymbol{\sigma}(\boldsymbol{x}_t) = \begin{bmatrix} b\sigma(x_{1,t}, x_{2,t}) \\ -c\sigma(x_{1,t}, x_{2,t}) \end{bmatrix}. \tag{42}$$

Clearly combinations of Case 1 and Case 2 are relevant, unfortunately explicit Lamperti type transformations are impossible (see Gilioli et al. (2008) for MCMC based estimation in such an example).

4.1 Example 2: Stochastic birth and death processes

The transformation is a direct application of Theorem 2 and for $\gamma_i \neq 1$ the transformation is given by

$$\psi^i(x_{i,t}) = \int \frac{d\xi}{\sigma_i \xi^{\gamma_i}}\bigg|_{\xi=x_{i,t}} = \frac{x_{i,t}^{1-\gamma_i}}{\sigma_i(1-\gamma_i)} \quad \Rightarrow \quad x_{i,t} = [\sigma_i(1-\gamma_i)z_{i,t}]^{\frac{1}{1-\gamma_i}},$$

For $\gamma_i = 1$ the transformation is the log-transform, for $\gamma_i > 1$ the linear growth condition for diffusions processes is not fulfilled and this case will therefore be disregarded.

For $\gamma_i < 1$, the state space for $\boldsymbol{z}_t$ is equal to the state space of $\boldsymbol{x}_t$, and the transformed system is given by (Appendix A.1)

$$d\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma_1}x_1^{1-\gamma_1}(a-bx_2) - \frac{1}{2}\gamma_1 x_1^{\gamma_1-1}\sigma_1(1+r^2) \\ \frac{1}{\sigma_1}x_2^{1-\gamma_2}(cx_1-d) - \frac{1}{2}\gamma_2 x_2^{\gamma_2-1}\sigma_2(1+r^2) \end{bmatrix} dt + \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} d\boldsymbol{w}. \tag{43}$$

In the limit $x_{1,t} \to 0$, the drift term in (43) approach $-\infty$, and there is a positive probability of $\boldsymbol{z}_t$ leaving the state space. Even though it might be possible to find the smoothen state by the EKF in specific situations, it will be difficult to estimate parameters in this model.

For $\gamma_i = 1$ the transformed system is given by

$$d\begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma_1}(a-bx_{2,t}) - \frac{1}{2}\sigma_1(1+r^2) \\ \frac{1}{\sigma_2}(cx_{1,t}-d) - \frac{1}{2}\sigma_2(1+r^2) \end{bmatrix} dt + \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} d\boldsymbol{w}. \tag{44}$$

The state space for $\boldsymbol{z}_t$ is $\mathbb{R}^2$ and as $\boldsymbol{x}_t$ approaches zero, we get a system with constant drift and additive noise, implying that $\boldsymbol{z}_t$ does not explode in finite time almost surely ($a.s.$).

4.1.1 Constant of motion

The deterministic Lotka-Voltarre system admits a constant of motion given by (Klebaner, 2005)

$$K_t = x_{1,t}^d x_{2,t}^a e^{-(cx_{1,t}+bx_{2,t})}, \tag{45}$$

implying that for any given initial conditions, K_0, $dK_t/dt = 0 \ \forall \ t$. The maximum of K_t is obtained at the stationary point and the log-transform of K_t is

given by

$$\log(K_t) = d\log(x_{1,t}) + a\log(x_{2,t}) - (cx_{1,t} + bx_{2,t}) \quad (46)$$
$$= d \cdot z_{1,t} + az_{2,t} - (ce^{z_{1,t}} + be^{z_{2,t}}), \quad (47)$$

for the stochastic system analysed here, the Itô process for $z_t^K = \log(K_t)$ can be derived by applying Itô's Lemma (see Appendix B.1)

$$dz_t^K = -\frac{1}{2}(1+r^2)(d\sigma_1^2 + a\sigma_2^2)dt + \tilde{\boldsymbol{\sigma}}(x_{1,t}, x_{2,t})d\boldsymbol{w}_t, \quad (48)$$

with

$$\tilde{\sigma}_1(x_{1,t}, x_{2,t}) = \sigma_1(d - cx_{1,t}) + r\sigma_2(a - bx_{2,t}) \quad (49)$$
$$\tilde{\sigma}_2(x_{1,t}, x_{2,t}) = r\sigma_1(d - cx_{1,t}) + \sigma_2(a - bx_{2,t}), \quad (50)$$

implying that there is a negative drift in the "constant of motion", further the process will be dominated by the drift term when the process is close to the stationary points, this negative drift is referred to as the likely path to extinction by Klebaner (2005).

The negative drift of the "constant of motion" depends on the diffusion parameters (σ_1, σ_2, r), and to obtain a system which does not explode on reasonable time scales (i.e. we want to see several periods) the leading diffusion parameters (σ_1, σ_2), should be small compared to the parameters controlling the deterministic dynamics of the model.

4.1.2 Simulation and estimation

The process was simulated 100 times using the Euler approximation (Kloeden & Platen, 1999) with $\Delta t = 10^{-7}$, and the parameters given in Table 1. Figure 2 show three of the simulated path in the phase plane, $\boldsymbol{z}_t^1$ and $\boldsymbol{z}_t^3$ are chosen to span the variation in the obtained simulations (in terms of z_t^K), $\boldsymbol{z}_t^2$ is the trajectory with the smallest variation around $E[z_t^K]$, which actually coincide with the median of the obtained "constants of motion" at time $t = 10$. Figure 3 present time series plots of the same data.

For each of the independent simulations the parameters were estimated and Table 1 present a summary of the estimation results. The likelihood estimation provide a parameter estimate and an estimate of the parameter variance.

Two tests are performed to evaluate the quality of the estimation results. Under the null hypothesis ($s^2 = \bar{s^2}_{est}$), the test parameter $Z_F = s^2/\bar{s^2}_{est} \sim F(99, 100)$.

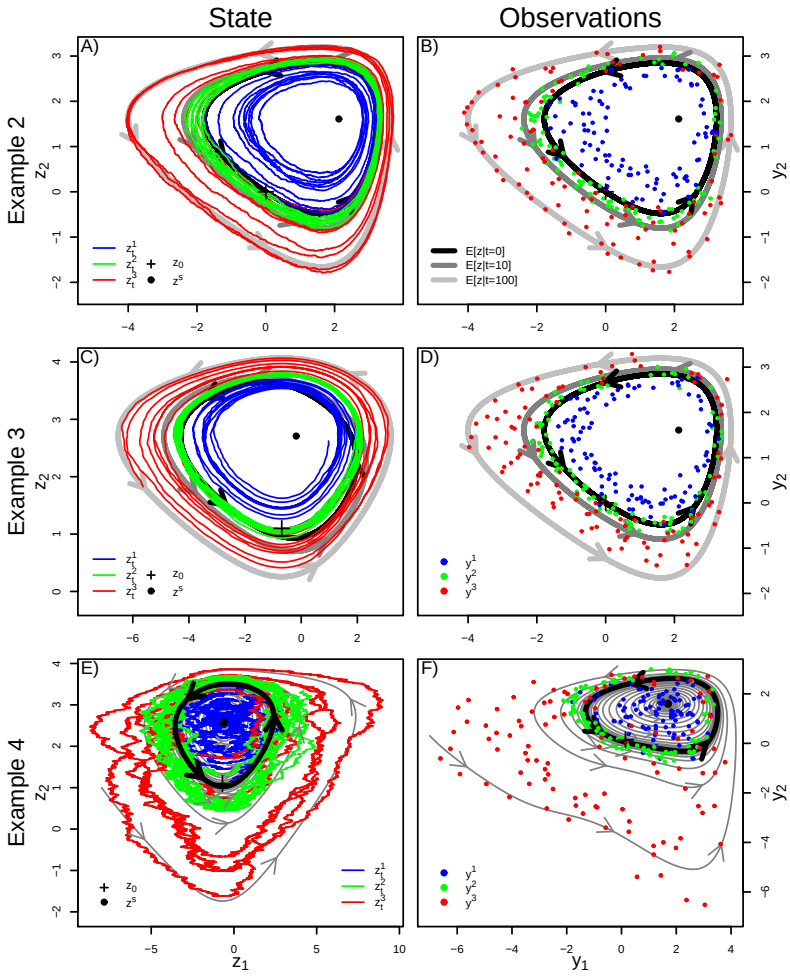

Figure 2: Simulation results of the simulation model presented in Section 4.1-4.3. The closed gray orbits in panel A-D represents the orbits corresponding to the expected value of the logarithm of the constant of motion in Example 2 at times 0, 10, 100. In E) and F) the closed orbit is the stationary limit cycle of the model Eq. (64), and gray lines represents transient behaviour. The shown simulations are chosen to span the variation of the observed dynamics.

The null hypothesis for the parameter estimates is $\bar{\theta} = \theta_{sim}$ and under the null hypothesis we have $Z_t = (\bar{\theta} - \theta_{sim})/(s\sqrt{N}) \sim t(99)$ (with N=100). Generally the mean parameter values are capture well (Table 1), while the parameter variances are not in the centre of the F-distribution. In particular the variance of

the diffusion parameters is not well described, and the variance of the correlation coefficient (r) is very large. One problem might be that the correlation r is restricted to the interval $(-1, 1)$ in the estimation, and there are many estimation results close to the upper bound. This will effect the quality of the estimation and in particular the estimated variances. The general picture is however that the mean values are well captured by the estimation.

Table 1: Summary of estimation results based on 100 independent simulations of Example 2 (Section 4.1). θ_{sim} and $\bar{\theta}$ are true and avrage estimated paramter values, s^2 and $\bar{s^2}_{est}$ is the empirical and estimated variances of the parameters, F-test and t-test are quantile in the F and student-t disribution for test of $s^2 = \bar{s^2}_{est}$ and $\theta_{sim} = \bar{\theta}$.

	θ_{sim}	$\bar{\theta}$	s^2	$\bar{s^2}_{est}$	F-test	t-test
a	10.0	9.996	0.0117	0.0115	0.531	0.502
b	2.0	2.002	0.0017	0.0017	0.522	0.498
c	1.0	1.001	0.0004	0.0005	0.294	0.498
d	5.0	4.998	0.0046	0.0043	0.637	0.501
σ_1	0.1	0.079	0.0013	0.0023	0.004	0.523
σ_2	0.1	0.082	0.0006	0.0013	<0.001	0.530
r	0.5	0.588	0.2925	0.4871	0.006	0.494
s_1	0.1	0.099	0.0001	0.0001	0.217	0.503
s_2	0.1	0.098	0.0001	0.0001	0.252	0.510
$z_{1,0}$	0.0	-0.002	0.0027	0.0041	0.018	0.502
$z_{2,0}$	0.0	0.004	0.0037	0.0046	0.133	0.497

4.2 Example 3: Stochastic predation

Consider the diffusion (42) and the linear transformation

$$s_t = cx_{1,t} + bx_{2,t} \quad \Rightarrow x_{2,t} = \frac{s_t - cx_{1,t}}{b}, \tag{51}$$

s_t can be consider as the total biomass in the system measured as in units of herbivores, the dynamics of $[x_{1,t}, s_t]^T$ is governed by the SDE

$$d\begin{bmatrix} x_{1,t} \\ s_t \end{bmatrix} = \begin{bmatrix} ax_{1,t} - bx_{1,t}x_{2,t} \\ acx_{1,t} - bdx_{2,t} \end{bmatrix} dt + \begin{bmatrix} b\sigma(x_{1,t}, x_{2,t}) \\ 0 \end{bmatrix} dw \tag{52}$$

$$= \begin{bmatrix} (ax_{1,t} - x_{1,t}(s_t - cx_{1,t})) \\ c(a+d)x_{1,t} - d \cdot s_t \end{bmatrix} dt + \begin{bmatrix} b\sigma\left(x_{1,t}, \frac{s_t - cx_{1,t}}{b}\right) \\ 0 \end{bmatrix} dw. \tag{53}$$

A natural choice of the diffusion process would be $\sigma(x_1, x_2)) = \sigma_0 x_1^\alpha x_2^\beta$. Even though the Lamperti transform exist in explicit form, with an explicit inverse, for $\alpha = \beta = \frac{1}{2}$ we encounter the same problem as for $\gamma_i < 1$ in the previous example.

If α and β are both chosen as 1, we get the constant of motion (Appendix A.2)

$$dz_t^K = -\frac{1}{2}\sigma_0^2 \left(ac^2 x_{1,t}^2 + db^2 x_{2,t}^2\right) dt + \sigma_0(cax_{1,t} - bdx_{2,t})dw_t. \tag{54}$$

The drift is proportional to the squared states, which might give a very fast drift towards extinction. To avoid this behaviour of the transition probabilities we suggest the following formulation of the diffusion (see Gilioli et al. (2008) for an alternative way to kill the diffusion as the state variable become large)

$$\sigma(x_{1,t}, x_{2,t}) = \sigma_0 \frac{x_{1,t}x_{2,t}}{cx_{1,t} + bx_{2,t}} = \frac{\sigma_0}{b} \frac{x_{1,t}(s_t - cx_{1,t})}{s_t}. \tag{55}$$

In this case the constant of motion z_t^K is determined by the SDE (Appendix B.3)

$$dz_t^K = -\frac{\sigma_0^2}{2} \frac{ac^2 x_{1,t}^2 + b^2 dx_{2,t}^2}{(cx_{1,t} + bx_{2,t})^2} dt - \sigma_0 \frac{acx_{1,t} - bdx_{2,t}}{cx_{1,t} + bx_{2,t}} dw_t, \tag{56}$$

the extreme behaviour of the drift term is now avoided, and $x_{1,t}$ is determined by the process

$$dx_{1,t} = (ax_{1,t} - x_{1,t}(s_t - cx_{1,t}))dt + \sigma_0 \frac{x_{1,t}(s_t - x_{1,t})}{s_t} dw. \tag{57}$$

Choosing the noise term (56) will lead to the Lamperti-transform

$$z_{1,t} = \psi(x_{1,t}, s_t) = \log\left(\frac{x_{1,t}}{s_t - cx_{1,t}}\right) \quad \Rightarrow \quad x_{1,t} = \frac{s_t}{c + e^{-z_{1,t}}} \tag{58}$$

If we consider the log-transformed $(z_{2,t})$ of s_t, the dynamics of the transformed system is given by (Appendix A.2)

$$d\begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} = \begin{bmatrix} a+d-s_t+\sigma_0^2\left(c\frac{x_{1,t}}{s_t}-\frac{1}{2}\right) \\ -d+c(a+d)\frac{x_{1,t}}{s_t} \end{bmatrix} dt + \begin{bmatrix} \sigma_0 \\ 0 \end{bmatrix} dw \tag{59}$$

$$= \begin{bmatrix} a+d-e^{z_{2,t}}+\sigma_0^2\left(\frac{c}{c+e^{-z_{1,t}}}-\frac{1}{2}\right) \\ -d+\frac{c(a+d)}{c+e^{-z_{1,t}}} \end{bmatrix} dt + \begin{bmatrix} \sigma_0 \\ 0 \end{bmatrix} dw, \tag{60}$$

and the state space of $[z_{1,t}, z_{2,t}]$ is $\mathbb{R}^2$.

4.2.1 Simulation and estimation

The process was again simulated 100 times with the parameters given in Table 2 and phase plane plots are given in Figure 2, while time series plot are given in Figure 3, the shown realisations are again chosen to span the variation in the realised simulation path. The parameters of each of the independent simulations were again estimated and the test summary is given in Table 2. The problems of the estimation procedure in the previous example is not present, both the parameter variance and the parameter estimates are well determined by the estimation procedure. The only value which does not have good test statistics is the variance of the initial value of $z_{1,t}$.

Table 2: Summary of estimation results based on 100 independent simulations of Example 3 (Section 4.2). θ_{sim} and $\bar{\theta}$ are true and avrage estimated paramter values, s^2 and $\bar{s^2}_{est}$ is the empirical and estimated variances of the parameters, F-test and t-test are quantiles in the F and student-t disribution for test of $s^2 = \bar{s^2}_{est}$ and $\theta_{sim} = \bar{\theta}$.

	θ_{sim}	$\bar{\theta}$	s^2	$\bar{s^2}_{est}$	F-test	t-test
a	10.000	10.007	0.0083	0.0096	0.230	0.497
b	2.000	2.003	0.0014	0.0016	0.280	0.496
c	1.000	1.002	0.0004	0.0004	0.339	0.495
d	5.000	5.002	0.0023	0.0024	0.447	0.499
σ_0	0.200	0.183	0.0014	0.0013	0.663	0.519
s_1	0.100	0.099	0.0001	0.0001	0.314	0.507
s_2	0.100	0.098	0.0001	0.0001	0.561	0.508
$z_{1,0}$	-0.693	-0.710	0.0070	0.0097	0.057	0.508
$z_{2,0}$	0.955	1.101	0.0018	0.0017	0.611	0.366

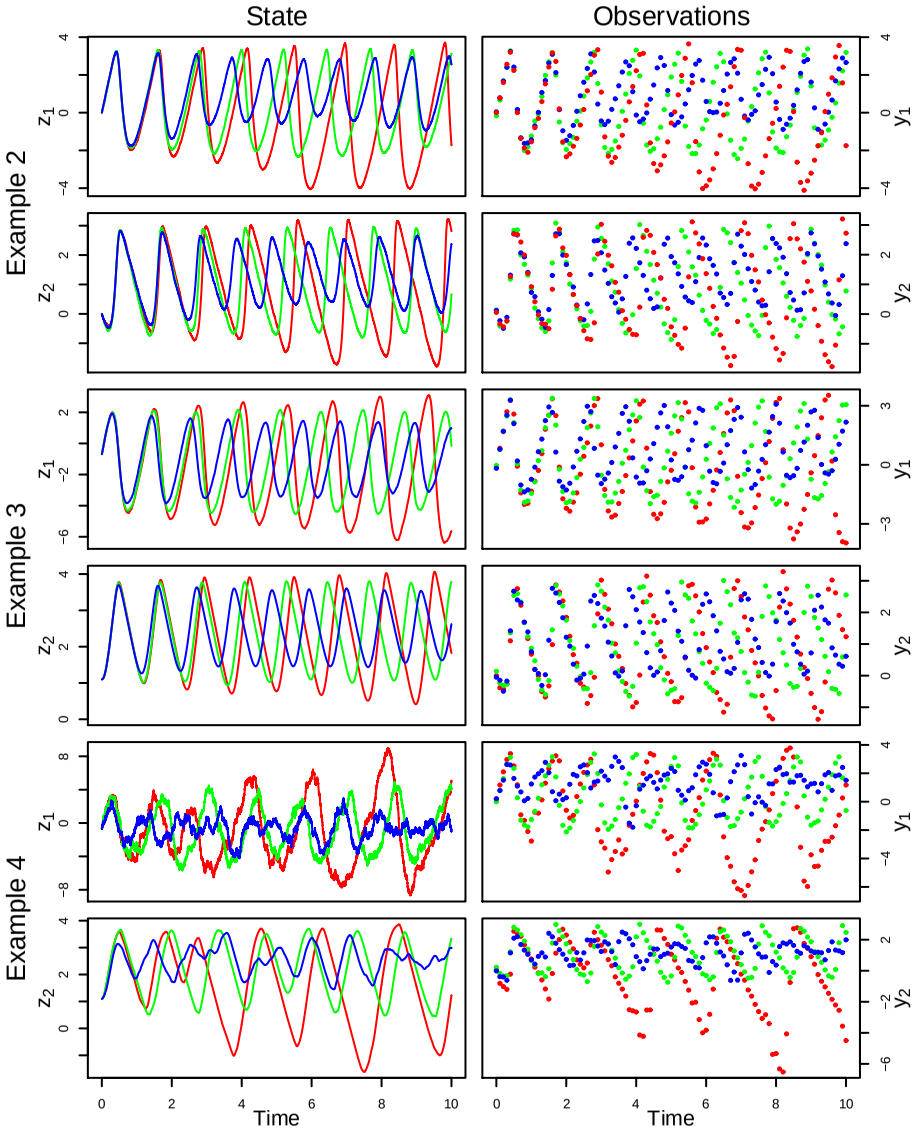

Figure 3: Time series plot of the data presented in Figure 2.

4.3 Example 4: A noise dominated system

The dynamics the two previous examples are dominated by the drift term, and as we have seen the dynamics of the deterministic system and the diffusion process might be very different. The divergence towards extension relies on the relative (to the drift parameters) size of the diffusion parameters (σ_0, σ_1, and σ_2), it is therefore difficult to use these examples with large diffusion parameters. To evaluate the estimation methodology in a system with large values of the diffusion term we use a modified drift term which compensate for the drift created by the diffusion. The construction is a simple modification of the Lotka-Voltarre model (38),

$$d\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} ax_{1,t}\left(1-\frac{x_{1,t}}{\kappa}\right) - bx_{1,t}x_{2,t} \\ cx_{1,t}x_{2,t} - dx_{2,t} \end{bmatrix} dt + \boldsymbol{\sigma}(\boldsymbol{x}_t)d\boldsymbol{w}, \tag{61}$$

the construction imply that we have imposed a maximum carrying capacity (κ) in the prey population, and the drift of (61) is equivalent to the construction in Gilioli et al. (2008). Further Zhu & Yin (2009) show that introducing carrying capacities in both states will lead to systems where the time to explosion of any of the state variables is equal ∞ *a.s.*. We can therefore expect the dynamics of the system to be more stable than for the previous example and the system is therefore better suited for simulations with large values of the diffusion parameter.

The dynamics of the "constant of motion", with the diffusion term as in Example 3, is determined by (Appendix B.4)

$$\begin{aligned} dz_t^K = &\left(\frac{ax_1}{\kappa}(cx_1-d) - \frac{\sigma_0^2}{2}\frac{ac^2x_1^2 + b^2dx_2^2}{(cx_1+bx_2)^2}\right)dt \\ &- \sigma_0\frac{acx_1 - bdx_2}{cx_1+bx_2}dw_t, \end{aligned} \tag{62}$$

when x_1 is large there is a positive drift of z_t^K, which will compensate for the negative drift created by the diffusion term. Clearly z_t^K is not a constant of motion for the deterministic skeleton of (61).

If κ is large the deterministic skeleton have a stable fix point at $(x_1, x_2) = \left(\frac{c}{d}, \frac{a}{b}\left(1-\frac{d}{c\kappa}\right)\right)$. The diffusion imposed an outward drift in the original system, and for a range of values for the diffusion the system (61) will have a stable limit cycle in the Lamperti transformed domain. The purpose of this study is not an in depth analysis of the system (61) and we will simply present one choice of κ and σ_0 in the system (61) with stochastic predation. The Lamperti transformed

system is given by (Appendix A.3)

$$d\begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} = \begin{bmatrix} a\left(1-\frac{x_{1,t}}{\kappa}\right)+d-s_t-\frac{\sigma_0^2(s_t-2x_{1,t})}{2s_t} \\ -d+c(a+d)\frac{x_{1,t}}{s_t}-c\frac{a}{\kappa}\frac{x_{1,t}^2}{s_t} \end{bmatrix} dt + \begin{bmatrix} \sigma_0 \\ 0 \end{bmatrix} dw \tag{63}$$

$$= \begin{bmatrix} a+d-e^{z_{2,t}}-\frac{a}{\kappa}\frac{e^{z_{2,t}}}{c+e^{-z_{1,t}}}+\sigma_0^2\left(\frac{c}{c+e^{-z_{1,t}}}-\frac{\sigma_0^2}{2}\right) \\ -d+\frac{c(a+d)}{c+e^{-z_{1,t}}}-c\frac{a}{\kappa}\frac{e^{z_{2,t}}}{(c+e^{-z_{1,t}})^2} \end{bmatrix} dt$$

$$+ \begin{bmatrix} \sigma_0 \\ 0 \end{bmatrix} dw. \tag{64}$$

4.3.1 Simulation and testing

The system was simulated 100 times with the parameters given in Table 3, and the same simulation setting as the previous examples. The phase-plane plots of the simulation (Figure 2) show strong stochastic behaviour compared with the previous examples, the shown realisations are chosen to span the variations in the realised processes, i.e. the realisations with the smallest, the median, and the largest mean distance to the stationary point (in the transformed domain) are shown. The time-series plot are given in Figure 3, and the separation in a noisy component $z_{1,t}$ and a "deterministic" component $z_{2,t}$ is very clear in the time domain.

Table 3: Summary of estimation results based on 100 independent simulations of Example 4 (Section 4.3). θ_{sim} and $\bar{\theta}$ are true and avrage estimated paramter values, s^2 and $\bar{s^2}_{est}$ is the empirical and estimated variances of the parameters, F-test and t-test are quantile in the F and student-t disribution for test of $s^2 = \bar{s^2}_{est}$ and $\theta_{sim} = \bar{\theta}$.

	θ_{sim}	$\bar{\theta}$	s^2	$\bar{s^2}_{est}$	F-test	t-test
a	10.000	10.005	0.1040	0.0638	0.992	0.499
b	2.000	2.046	0.0141	0.0072	0.999	0.485
c	1.000	1.027	0.0028	0.0026	0.661	0.479
d	5.000	4.943	0.0084	0.0061	0.943	0.525
κ	50.000	47.440	14.2465	11.7214	0.833	0.527
σ_0	2.500	2.396	0.0237	0.0242	0.461	0.527
s_1	0.100	0.127	0.0009	0.0004	>0.999	0.465
s_2	0.100	0.113	0.0002	0.0002	0.916	0.467
$z_{1,0}$	-0.693	-0.447	0.2995	0.8522	<0.001	0.482
$z_{2,0}$	0.955	1.097	0.0081	0.0306	<0.001	0.437

The estimation results are summarised in Table 3, the mean of the parameter values are still captured quite well with test statistics in the centre of the t-distribution, the performance of the variance estimates are however generally quite poor, and the hypothesis of $s^2 = \bar{s^2}_{sim}$ is rejected for most of the estimates.

5 Discussion and conclusion

The corrected results from Nielsen & Madsen (2001) provide a general formulation of multivariate Lamperti transformations for a restricted class of multivariate SDEs. Example 1 (Section 3.1) illustrates that we can expect good performance of state estimates by the EKF, also when the diffusion dominates the system, at the same time the example illustrate that even though the state estimates by the EKF are close the mode of the true posterior distribution, the shape of the distribution might not be captured well when the diffusion parameter is large.

The parameter estimation in the Lotka-Volterra examples (Section 4) generally perform well when regarding the mean value of the parameter estimates. When diffusion is dominant in the system, the estimation procedure do not give reliable estimates of the parameter variance. CTSM use the inverse Hessian of the log-likelihood, at the optimal parameter values to form an estimate of the parameter covariance matrix. The conclusions therefore fit well with the conclusion from Example 1, where the estimated mode was close to the true mode, while the shape of the distribution was not captured well. The conclusions above clearly depends heavily on the distance between sampling points and as sampling distance increase, the reliability of the parameter estimates and the parameter variance estimates will decrease.

References

Aït-Sahalia Y. (2008) Closed-form likelihood expansions for multivariate diffusions *The Annals of Statistics*, **32**, (2), 906-937.

Baadsgaard M., Nielsen J.N., Spliid H., Madsen H., and Preisel M. (1997) Estimation in stochastic differential equations with state dependent diffusion term *SYSID '97 - 11th IFAC symposium on system identification, IFAC.*

Gilioli G., Pasquali S., and Ruggeri F. (2008) Bayesian Inference for Functional Response in a Stochastic Predator-Prey System. *Bulletin of Mathematical Biology*, **70**, 358-381.

Givon D., Stinis P. and Weare J. (2009) Variance reduction for Particle filters of Systems With Time Scale Separation. *IEEE Transactions on Signal Processing*, **57**, (2), 424-435.

Iacus S.M. (2008) Simulation and Inference for Stochastic Differential Equations - With R Examples *Springer Series in Statistics.*

Jazwinski A.H. (1970) Stochastic Processes and Filtering Theory, *Academic Press*, New York.

Kalman R.E. (1960) A New Approach to Linear Filtering and Prediction Problems, *Transactions of the ASME-Journal of Basic Engineering, Series D*, **82**, 35-45.

Klebaner F.C. (2005) Introduction to Stochastic Calculus with Applications. *Imperial College Press.*

Kloden, P., and Platen, E. (1999) Numerical solutions of Stochastic Differential Equations. *Springer-Verlag, Berlin.*

Kristensen, N. R., and Madsen, H. (2003) Continuous time stochastic modelling - CTSM 2.3 - Mathematics Guide. *Technical University of Denmark.*

Kristensen N.R., Madsen H., and Jørgensen S.B. (2004) Parameter estimation in stochastic grey-box models, *Automatica*, **40**, 225-237.

Luschgy H. and Pagés G. (2006) Functional quantization of a class of Brownian diffusions: A constructive approach *Stochastic Processes and their Applications* **116**, *310-336.*

Møller J.K., and Madsen H. (2010) From Level Dependent Diffusion to Constant Diffusion in Stochastic Differential Equations by the Lamperti Transform. *IMM-technical report 2010-16, DTU-Informatics, Technical University of Denmark.*

Nielsen J.N., and Madsen H. (2001) Applying the EKF to stochastic differential equations with level effects. *Automatica,* **37***, 107-112.*

Simon D. (2006) Optimal state estimation- Kalman, H_∞, and nonlinear approaches. *John Wiley & Sons, Inc.*

Zhu C. and Yin G. (2009) On hybrid competitive Lokta-Volterra ecosystems *Nonlinear Analysis*, **71**, 1370-1379.

Øksendal B. (2003) Stochastic Differential Equations - An Introduction with Applications, Sixth edition. *Springer*, Berlin.

A Derivations transformed system equations

A.1 Equation (43)

We start out by calculating the derivatives of the transformation

$$\psi^i(x_i) = \frac{x_i^{1-\gamma_i}}{\sigma_i(1-\gamma_i)} \tag{A.1}$$

the partial derivatives are given by

$$\psi^i_{x_i} = \frac{1}{\sigma_i} x_i^{-\gamma_i}; \quad \psi^i_{x_i x_j} = \begin{cases} -\frac{1}{\sigma_i} x_i^{-1-\gamma_i} & i = j \\ 0 & i \neq j. \end{cases} \tag{A.2}$$

Inserting in the Lokta Volterra system we get

$$\begin{aligned} d\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} &= \begin{bmatrix} (ax_1 - bx_1x_2)\psi^1_{x_1} + \frac{1}{2}\psi_{x_1x_1}\sigma_1^2 x_1^{2\gamma_1}(1+r^2) \\ (cx_1x_2 - dx_2)\psi^2_{x_2} + \frac{1}{2}\psi_{x_2x_2}\sigma_2^2 x_2^{2\gamma_2}(1+r^2) \end{bmatrix} dt \\ &\quad + \begin{bmatrix} \sigma_1 x_1^{\gamma_1}\psi^1_{x_1} & 0 \\ 0 & \sigma_2 x_2^{\gamma_2}\psi^2_{x_2} \end{bmatrix} \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} d\boldsymbol{w} \qquad \text{(A.3)} \\ &= \begin{bmatrix} \frac{1}{\sigma_1}(ax_1^{1-\gamma_1} - bx_1^{1-\gamma_1}x_2) - \frac{1}{2}\gamma_1\sigma_1 x_1^{\gamma_1-1}(1+r^2) \\ \frac{1}{\sigma_2}(cx_1x_2^{1-\gamma_2} - dx_2^{1-\gamma_2}) - \frac{1}{2}\gamma_2\sigma_2 x_2^{\gamma_2-1}(1+r^2) \end{bmatrix} dt \\ &\quad + \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} d\boldsymbol{w} \qquad \text{(A.4)} \\ &= \begin{bmatrix} \frac{x_1^{1-\gamma_1}}{\sigma_1}(a - bx_2) - \frac{1}{2}\gamma_1 x_1^{\gamma_1-1}\sigma_1(1+r^2) \\ \frac{x_2^{1-\gamma_2}}{\sigma_2}(cx_1 - d) - \frac{1}{2}\gamma_2 x_2^{\gamma_2-1}\sigma_2(1+r^2) \end{bmatrix} dt + \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} d\boldsymbol{w}. \qquad \text{(A.5)} \end{aligned}$$

Which was to be shown.

A.2 Equation (60)

With $z_1 = \psi(x_1, s)$ and

$$\psi(x_1, s) = \log\left(\frac{x_1}{s - cx_1}\right) \tag{A.6}$$

the needed partial derivatives are given by

$$\psi_{x_1} = \frac{s}{x_1(s-cx_1)}, \quad \psi_s = -\frac{1}{s-cx_1}, \quad \psi_{x_1x_1} = -\frac{s(s-2cx_1)}{x_1^2(s-cx_1)^2} \tag{A.7}$$

and the dynamics of z_1 is determined by (note that $(ds)^2 = dsdx_1 = 0$)

$$\begin{aligned} dz_1 =& dx_1\psi_{x_1} + ds\psi_s + \frac{1}{2}(dx_1)^2\psi_{x_1x_1} && \text{(A.8)}\\ =&(ax_1 - x_1(s-cx_1))\frac{s}{x_1(s-cx_1)}dt + \sigma_0\frac{x_1(s-cx_1)}{s}\psi_{x_1}dw \\ &- (c(a+d)x_1 - d\cdot s)\frac{1}{s-cx_1}dt - \frac{\sigma_0^2}{2}\frac{x_1^2(s-cx_1)^2}{s^2}\frac{s(s-2cx_1)}{x_1^2(s-cx_1)^2}dt \\ =&\left(\frac{as - c(a+d)x_1 + d\cdot s}{s-cx_1} - s - \frac{\sigma_0^2}{2}\frac{s-2cx_1}{s}\right)dt + \sigma_0 dw && \text{(A.9)}\\ =&\left(a + d - s - \sigma_0^2\left(\frac{1}{2} - c\frac{x_1}{s}\right)\right)dt + \sigma_0 dw && \text{(A.10)} \end{aligned}$$

which was to be shown.

A.3 Equation (64)

The dynamics of s_t is determined by the Itô-SDE

$$\begin{aligned} ds_t =& cax_1\left(1 - \frac{x_1}{\kappa}\right) - d(s - cx_1) && \text{(A.11)}\\ =& cx_1\left[a\left(1 - \frac{x_1}{\kappa}\right) + d\right] - ds, && \text{(A.12)} \end{aligned}$$

inserting $x_1 = \frac{s}{c+e^{-z_1}}$ and $s = e^{z_2}$ gives the equation for z_2. To find dz_1 we insert the partial derivatives from Section A.2 in

$$\begin{aligned} dz_1 =& dx_1\psi_{x_1} + ds\psi_s + \frac{1}{2}(dx_1)^2\psi_{x_1x_1} && \text{(A.13)}\\ =& \left[\frac{\left(ax_1\left(1-\frac{x_1}{\kappa}\right) - x_1(s-cx_1)\right)s}{x_1(s-cx_1)} - \frac{cx_1\left(a\left(1-\frac{x_1}{\kappa}\right)+d\right) - ds}{s-cx_1}\right]dt \\ &+ \frac{\sigma_0^2}{2}\frac{(2cx_1 - s)(s-cx_1)^2}{s(s-cx_1)^2}dt + \sigma_0 dw_t && \text{(A.14)}\\ =& \left[-s + \frac{a\left(1-\frac{x_1}{\kappa}\right)s - cx_1\left(a\left(1-\frac{x_1}{\kappa}\right) - d\right) + ds}{s-cx_1}\right]dt \\ &+ \sigma_0^2\left(c\frac{x_1}{s} - \frac{1}{2}\right)dt + \sigma_0 dw_t && \text{(A.15)}\\ =& \left[-s + \frac{a(s-cx_1)\left(1-\frac{x_1}{\kappa}\right) + d(s-cx_1)}{s-cx_1} + \sigma_0^2\left(c\frac{x_1}{s} - \frac{1}{2}\right)\right]dt \\ &+ \sigma_0 dw_t && \text{(A.16)} \end{aligned}$$

$$= \left[-s + a\left(1 - \frac{x_1}{\kappa}\right) + d + \sigma_0^2 \left(c\frac{x_1}{s} - \frac{1}{2} \right) \right] dt + \sigma_0 dw_t \tag{A.17}$$

substituting with z_1 and z_2 gives the desired result.

B Derivations stochastic "constants" of motion

B.1 Derivation of Eq. (48)

With

$$\psi^K(z_1, z_2) = dz_1 + az_2 - ce^{z_1} - be^{z_2}, \tag{B.18}$$

the partial derivatives are given by

$$\psi^K_{z_1} = d - ce^{z_1}; \qquad \psi^K_{z_1 z_1} = -ce^{z_1} \tag{B.19}$$

$$\psi^K_{z_2} = a - be^{z_2}; \qquad \psi^K_{z_2 z_2} = -be^{z_2} \tag{B.20}$$

$$\psi^K_{z_1 z_2} = 0 \tag{B.21}$$

with $z_t^K = \psi(z_{1,t}, z_{2,t})$ we get

$$\begin{aligned} dz_t^K = & \left[\left(a - \frac{1}{2}\sigma_1^2(1+r^2) - be^{z_2} \right) dt + \sigma_1 dw_{1,t} + r\sigma_1 dw_{2,t} \right] \psi^K_{z_1} \\ & + \left[\left(ce^{z_1} - \frac{1}{2}\sigma_2^2(1+r^2) - d \right) dt + r\sigma_2 dw_{1,t} + \sigma_2 dw_{2,t} \right] \psi^K_{z_2} \\ & + \left[\sigma_1^2(1+r^2) \right] dt \psi^K_{z_1 z_1} + \left[\sigma_2^2(1+r^2) \right] dt \psi^K_{z_2 z_2} \end{aligned} \tag{B.22}$$

$$\begin{aligned} = & \left[(a - be^{z_2})\, \psi^K_{z_1} + (ce^{z_1} - d)\, \psi^K_{z_2} \right] dt - \\ & \frac{1}{2}(1+r^2) \left[\sigma_1^2(\psi^K_{z_1} - \psi^K_{z_1 z_1}) + \sigma_2^2(\psi^K_{z_2} - \psi^K_{z_2 z_2}) \right] dt + \\ & (\sigma_1 \psi^K_{z_1} + r\sigma_2 \psi^K_{z_2}) dw_{1,t} + (r\sigma_1 \psi^K_{z_1} + \sigma_2 \psi^K_{z_2}) dw_{2,t} \end{aligned} \tag{B.23}$$

$$\begin{aligned} = & \left[\psi^K_{z_2} \psi^K_{z_1} + \left(-\psi^K_{z_1} \right) \psi^K_{z_2} \right] dt - \frac{1}{2}(1+r^2) \left[d\sigma_1^2 + a\sigma_2^2 \right] dt + \\ & (\sigma_1 \psi^K_{z_1} + r\sigma_2 \psi^K_{z_2}) dw_{1,t} + (r\sigma_1 \psi^K_{z_1} + \sigma_2 \psi^K_{z_2}) dw_{2,t} \end{aligned} \tag{B.24}$$

$$\begin{aligned} = & -\frac{1}{2}(1+r^2) \left[d\sigma_1^2 + a\sigma_2^2 \right] dt + \\ & (\sigma_1 \psi^K_{z_1} + r\sigma_2 \psi^K_{z_2}) dw_{1,t} + (r\sigma_1 \psi^K_{z_1} + \sigma_2 \psi^K_{z_2}) dw_{2,t}, \end{aligned} \tag{B.25}$$

replacing the partial derivatives of ψ by their functions of z_i, and z_i by $\log(x_i)$ gives the desired results.

B.2 Equation (54)

When $\alpha = \beta = 1$ the dynamics of the log-transformed system (with $z_i = \log(x_i)$) is governed by the SDE

$$d\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} a - bx_2 - \frac{1}{2}\sigma_0^2 b^2 x_2^2 \\ cx_1 - d - \frac{1}{2}\sigma_0^2 c^2 x_1^2 \end{bmatrix} dt + \sigma_0 \begin{bmatrix} bx_2 \\ -cx_1 \end{bmatrix} d\boldsymbol{w}_t. \tag{B.26}$$

Proceeding as in Section B.1 we get

$$\begin{aligned} dz_t^K = &\left[(a - be^{z_2})\,\psi_{z_1}^K + (ce^{z_1} - d)\,\psi_{z_2}^K\right] dt - \\ &\frac{1}{2}\sigma_0^2 \left[b^2 e^{2z_2}(\psi_{z_1}^K - \psi_{z_1 z_1}^K) + c^2 e^{2z_1}(\psi_{z_2}^K - \psi_{z_2 z_2}^K)\right] dt + \\ &\sigma_0 \left(be^{z_2}\psi_{z_1}^K - ce^{z_1}\psi_{z_2}^K\right) dw_t \qquad \text{(B.27)} \\ = &-\frac{1}{2}\sigma_0^2 \left[db^2 e^{2z_2} + ac^2 e^{2z_1}\right] dt \\ &+ \sigma_0 \left(be^{z_2}(d - ce^{z_1}) - ce^{z_1}(a - be^{z_2})\right) dw_t \qquad \text{(B.28)} \\ = &-\frac{1}{2}\sigma_0^2 \left[db^2 e^{2z_2} + ac^2 e^{2z_1}\right] dt + \sigma_0 \left(bde^{z_2} - ace^{z_1}\right) dw_t \qquad \text{(B.29)} \end{aligned}$$

inserting $x_i = e^{z_i}$ gives the desired result.

B.3 Equation (56)

The log-transformed system (with $z_i = \log(x_i)$) is governed by the SDE

$$d\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} a - bx_2 - \frac{1}{2}\sigma_0^2 \frac{x_2^2}{(cx_1 + bx_2)^2} \\ cx_1 - d - \frac{1}{2}\sigma_0^2 \frac{x_1^2}{(cx1 + bx2)^2} \end{bmatrix} dt + \frac{\sigma_0}{cx_1 + bx_2} \begin{bmatrix} bx_2 \\ -cx_1 \end{bmatrix} d\boldsymbol{w}_t. \tag{B.30}$$

Proceeding as in Section B.1 we get

$$\begin{aligned} dz_t^K = &\left[(a - be^{z_2})\,\psi_{z_1}^K + (ce^{z_1} - d)\,\psi_{z_2}^K\right] dt - \\ &\frac{\sigma_0^2}{2(ce^{z_1} + be^{z_2})^2} \left[b^2 e^{2z_2}(\psi_{z_1}^K - \psi_{z_1 z_1}^K) + c^2 e^{2z_1}(\psi_{z_2}^K - \psi_{z_2 z_2}^K)\right] dt + \\ &\frac{\sigma_0}{ce^{z_1} + be^{z_2}} \left(be^{z_2}\psi_{z_1}^K - ce^{z_1}\psi_{z_2}^K\right) dw_t \qquad \text{(B.31)} \\ = &-\frac{\sigma_0^2}{2(ce^{z_1} + be^{z_2})^2} \left[db^2 e^{2z_2} + ac^2 e^{2z_1}\right] dt \\ &+ \frac{\sigma_0}{ce^{z_1} + be^{z_2}} \left(be^{z_2}(d - ce^{z_1}) - ce^{z_1}(a - be^{z_2})\right) dw_t \qquad \text{(B.32)} \\ = &-\frac{\sigma_0^2 \left(db^2 e^{2z_2} + ac^2 e^{2z_1}\right)}{2(ce^{z_1} + be^{z_2})^2} dt + \frac{\sigma_0 \left(bde^{z_2} - ace^{z_1}\right)}{ce^{z_1} + be^{z_2}} dw_t \qquad \text{(B.33)} \end{aligned}$$

inserting $x_i = e^{z_i}$ gives the desired result.

B.4 Equation (62)

The log-transformed system (with $z_i = \log(x_i)$) is governed by the SDE

$$d\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} a\left(1-\frac{x_1}{\kappa}\right) - bx_2 - \frac{1}{2}\sigma_0^2 \frac{x_2^2}{(cx_1+bx_2)^2} \\ cx_1 - d - \frac{1}{2}\sigma_0^2 \frac{x_1^2}{(cx1+bx2)^2} \end{bmatrix} dt + \frac{\sigma_0}{cx_1+bx_2}\begin{bmatrix} bx_2 \\ -cx_1 \end{bmatrix} d\boldsymbol{w}_t, \tag{B.34}$$

and the dynamics of dz^K is determined by

$$\begin{aligned} dz_t^k = &\left[\left(a\left(1-\frac{e^{z_1}}{\kappa}\right) - be^{z_2}\right)\psi_{z_1}^K + \left(ce^{z_1}-d\right)\psi_{z_2}^K\right]dt - \\ &\frac{\sigma_0^2}{2(ce^{z_1}+be^{z_2})^2}\left[b^2e^{2z_2}(\psi_{z_1}^K - \psi_{z_1z_1}^K) + c^2e^{2z_1}(\psi_{z_2}^K - \psi_{z_2z_2}^K)\right]dt + \\ &\frac{\sigma_0}{ce^{z_1}+be^{z_2}}\left(be^{z_2}\psi_{z_1}^K - ce^{z_1}\psi_{z_2}^K\right)dw_t \qquad \text{(B.35)} \\ = &\left(-a\frac{e^{z_1}}{\kappa}\psi_{z_1}^K - \frac{\sigma_0^2\left(db^2e^{2z_2}+ac^2e^{2z_1}\right)}{2(ce^{z_1}+be^{z_2})^2}\right)dt \\ &+ \frac{\sigma_0\left(bde^{z_2}-ace^{z_1}\right)}{ce^{z_1}+be^{z_2}}dw_t \qquad \text{(B.36)} \\ = &\left(-a\frac{e^{z_1}}{\kappa}(d-ce^{z_1}) - \frac{\sigma_0^2\left(db^2e^{2z_2}+ac^2e^{2z_1}\right)}{2(ce^{z_1}+be^{z_2})^2}\right)dt \\ &+ \frac{\sigma_0\left(bde^{z_2}-ace^{z_1}\right)}{ce^{z_1}+be^{z_2}}dw_t \qquad \text{(B.37)} \end{aligned}$$

inserting $x_i = e^{z_i}$ gives the desired result.

Paper F

Structural Identification and Validation in Stochastic Differential Equation based Models - With application to a Marine Ecosystem NP-model

---

Authors:
J.K. Møller, H. Madsen, and J. Carstensen.

Submitted to:
Journal of the Royal Statistical Society, Series C, (2010).

Structural Identification and Validation in Stochastic Differential Equation based Models - With application to a Marine Ecosystem NP-model

Jan Kloppenborg Møller[1,2], Henrik Madsen[1] and Jacob Carstensen[2]

Abstact

Stochastic differential equations (SDEs) for ecosystem modelling have attracted increasing attention during recent years. The modelling has mostly been through simulation based experiments. Estimation of parameters in SDEs is, however, possible by combining Kalman filter and likelihood techniques. The resulting filter equations handle additive diffusion effectively, while state dependent diffusion is difficult to handle. In many cases it is however possible to transform the state-space to avoid state dependent descriptions. It is demonstrated how pure random walk hidden state formulation and state estimation of key parameters can generate data driven model formulations. The resulting models are based on short-term predictions and it is demonstrated how considerations on stationarity of the distribution and inspection of probabilistic properties of simulation results can generate further model improvements of simulation models. The proposed methodology is demonstrated using phytoplankton and nitrogen data from a Danish estuary covering a 16 years period (1988-2003). It is demonstrated how non-linear relationships between states can be identified by plotting the (random) production parameter as a function of the state variables and global radiation. Further improvements of both the drift and the diffusion term are achieved by comparing simulated densities and data.

Key Words: Stochastic differential equations, Maximum likelihood, Extended Kalman filter, Structural identification, Validation, Lamperti transform, Simulation performance, NP-models.

[1]DTU Informatics, Richard Pedersens Plads, Technical University of Denmark Building 321 DK-2800 Lyngby, Denmark.

[2]National Environmental Research Institute, Fredriksborgvej 399, DK-4000 Roskilde, Denmark.

1 Introduction

Stochastic differential equations (SDEs) are stochastic generalisations of ordinary differential equations (ODEs), where the differential increments are given a probabilistic interpretation (Øksendal (2003)). The theory of SDEs is in a mature state and the literature on theoretic properties of SDEs is bulk (e.g. Klebaner (2005), Karatzas and Shreve (1991)), and the use of SDEs is standard in mathematical finance.

SDEs has also proven useful in diverse fields such as pharmacokinetic (Tornøe et al. (2004)), engineering (Madsen at al. (1987)) and geolocation of fish (Pedersen et al. (2008)). These applications uses data to estimate parameters in a continuous-discrete time stochastic state space formulation, which allow a splitting of the noise processes into observation noise and system noise. In SDEs system noise is often referred to as diffusion, which describes the stochastic part of the state-space formulation.

The estimation in the present work is based on an implementation of the Extended Kalman Filter (EKF) (e.g. Jazwinski (1970)) and approximate likelihood estimation as presented in Kristensen et al. (2004a). The EKF allows for optimal state estimation, and through modelling parameters in the model as pure random walk hidden states it is possible to formulate data-driven hypotheses based on the reconstructed or smoothed state of the system (Kristensen et al. (2004b)).

The EKF filter approach is effective in handling additive (state independent) diffusion, however, such an assumption is in many cases a strong simplification of real life systems, that will not fulfil basic requirements of the system, such as positive states. The assumption also exclude a large class of well known diffusion processes (such as "Black and Scholes" type models, (Øksendal (2003)) and the Feller diffusion (Iacus (2008))). For one-dimensional diffusion processes this is effectively handled by transformation of the state-space (Baadsgaard et al. (1997)), the transformation is often referred to as the Lamperti transform (Iacus (2008)). For multivariate processes this is a more delicate matter (Aït-Sahalia (2008); Luschgy and Pagés (2006)), but for a restricted class of diffusion processes it can be handled by Itô's lemma, and a general multivariate formulation that allow for a Lamperti type transformation is presented.

The present work is a further development of the methodology presented in Kristensen et al. (2004b), in addition to the consideration based on the likelihood and reconstruction of the random walk hidden states, the structural development is based on considerations about the stationary solution and simulations

results. The proposed methodology is exemplified with a comprehensive study of a marine ecosystem.

Marine ecosystems represent very complex structures of coupled subprocesses of which each subprocess represent a detailed discipline in its own right, and the subprocesses interacts across a wide spectrum of space and time in a complicated manner, which ultimately determines the dynamics of the complete system.

Typically a model for a complex system is obtained by coupling deterministic sub-models together, where each sub-model describes a specific subprocess. The functional relations are therefore not based on the specific conditions observed at the study site. The output of the resulting model is compared to observations from the specific study-site and parameters are tuned to mimic observations in the ecosystem. An early and simple example of this approach is found in Fasham et al. (1990), a more resent and complex example is Bartell et al. (1999). The later example illustrates the profound complexity of ecosystem models.

The complexity of ecosystem models makes these especially useful for illustrating the methodology presented here, since the dynamics of the full system is not determined by the individual subprocess, but by the subprocesses *and* the way these are interconnected and working on different time scales.

Section 2 introduce the continuous-discrete time stochastic state-space formulation with emphasis on the transformations that enables estimation of system with state dependent noise. The proposed methodology is presented in Section 3, with the example constituting the main part of the article in Section 4. Finally the results from the example and some general implications of the proposed methodology are discussed in Section 5.

2 Continuous-discrete time stochastic state-space models

Stochastic differential equations are stochastic generalisations of ordinary differential equations in the sense that the deterministic skeleton of an SDE is an ODE. The continuous time state of the SDE is observed indirectly in discrete-time through the observation equation. This gives the continuous-discrete time stochastic state-space formulation

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{\theta}, t)dt + \boldsymbol{\sigma}(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{\theta}, t)d\boldsymbol{w}_t \tag{1}$$

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_{t_k}, \boldsymbol{u}_{t_k}, \boldsymbol{\theta}, \boldsymbol{e}_k, t_k), \tag{2}$$

where $t \in \mathbb{R}_0$ is time, t_k ($k \subset \mathbb{N}_0$) is the sample times, $\boldsymbol{x}_t \subseteq \chi \in \mathbb{R}^n$ is a vector of state variables belonging to the state-space (χ), $\boldsymbol{u}_t \in \mathbb{R}^r$ is a vector of inputs, $\boldsymbol{w}_t$ is the standard Brownian motion, $\boldsymbol{\theta} \in \mathbb{R}^p$ is a parameter vector, $\boldsymbol{f}(\cdot) \in \mathbb{R}^n$ is a vector function referred to as the drift term, $\boldsymbol{\sigma}(\cdot) \in \mathbb{R}^{m \times n}$ is a matrix function referred to as the diffusion term, $\boldsymbol{y}_k$ is the observation at time t_k, $\boldsymbol{h}(\cdot) \in \mathbb{R}^l$ is the observation function and $\boldsymbol{e}_k \in \mathbb{R}^l$ is a random observation error. Hence, the state-space formulation consist of the system equation (1), which describes the time-evolution of the states, and the observation equation (2), which describes the how the actual observations relates to the states.

The system equation (1) is a short-term notation for the integral interpretation, and in this context the Itô interpretation is used. Details on the formulation of SDEs and the general theory can be found in e.g. Øksendal (2003).

2.1 Parameter and state estimation

The estimation procedure employed here is based on the Extended Kalman Filter (EKF) and maximum likelihood estimation. A general account for the procedures can be found in Kristensen et al. (2004a), however the basic assumption is that the differential increments in Eq. (1) are Gaussian and that the observations are also Gaussian. For the filter equation to take on a sufficiently simple form to allow efficient implementation, the continuous-time stochastic state formulation and the discrete-time observation formulation is restricted to the form

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{\theta}, t)dt + \boldsymbol{\sigma}(\boldsymbol{u}_t, \boldsymbol{\theta}, t)d\boldsymbol{w}_t \tag{3}$$

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_{t_k}, \boldsymbol{u}_{t_k}, \boldsymbol{\theta}, t_k) + \boldsymbol{e}_k, \tag{4}$$

where $\boldsymbol{\sigma} \in \mathbb{R}^{n \times n}$ is a quadratic matrix function independent of the state, and $\boldsymbol{e}_k \in \mathbb{R}^l$ is a Gaussian random variable with zero mean and covariance $\boldsymbol{S}(\boldsymbol{u}_t, \boldsymbol{\theta}, t)$. All other terms are as explained above. The first restriction ($\boldsymbol{\sigma}$ quadratic) is not a real restriction since the estimation is based on the likelihood (weak solution) which, as a consequence of the fact that the Kolmogorov forward (Fokker-Planck) equation (Gard (1988)) only depends on $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$. The independence between the diffusion matrix and the state can to some extent be dealt with by transformation of the state-space (see below) to obtain a formulation where the diffusion is independent of the state. The last restriction (observation noise additive and Gaussian), is crucial for the EKF, real life observations are, however, often not Gaussian, but it is often possible to deal with this by transformations of the observations before estimation (e.g. Box-Cox transformations).

The approximate likelihood estimation is based on the assumption that the conditional density is Gaussian, and in this case the likelihood can be written

as

$$L(\boldsymbol{\theta};\mathcal{Y}_N) = p(\boldsymbol{y}_0|\boldsymbol{\theta}) \prod_{k=1}^{N} \frac{\exp\left(-\frac{1}{2}\boldsymbol{\epsilon}_k^T \boldsymbol{R}_{k|k-1}^{-1} \boldsymbol{\epsilon}_k\right)}{\sqrt{(2\pi)^l \det(\boldsymbol{R}_{k|k-1})}}, \tag{5}$$

where $\mathcal{Y}_N = \{\boldsymbol{y}_0, ..., \boldsymbol{y}_N\}$ are all observations up to time $T = t_N$, $p(\boldsymbol{y}_0|\boldsymbol{\theta})$ is the conditional (on the parameters) density of the first observation, the innovation covariance matrix is given by $\boldsymbol{R}_{k|k-1} = V\{\boldsymbol{y}_k|\mathcal{Y}_{k-1};\boldsymbol{\theta}\}$ and the innovation is given by $\boldsymbol{\epsilon}_k = \boldsymbol{y}_k - E\{\boldsymbol{y}_k|\mathcal{Y}_{k-1};\boldsymbol{\theta}\}$. While the restriction of the observation equation (4) is necessary for (5) to form a reasonable approximation, it is not sufficient, since the observation equation (4) consist of a function of the state $h(\cdot)$ *and* an additive Gaussian error. The assumption is therefore that the conditional density of $\boldsymbol{x}_t$ is approximately Gaussian (possibly after a transformation h). This is likely to hold when the sampling frequencies are fast (compared to the dynamics of the system), while there exist methods to verify this (Bak et al. (1999)), we will not be concerned with this issue in the present study, since the evaluation of the final model is with respect to long term simulations not short term predictions (see Section 3).

In addition to the parameter estimation provided by the maximum likelihood procedure the filtering procedure allows for state estimation to obtain the state reconstruction ($E[\boldsymbol{x}_t|\mathcal{Y}_t]$), and the smoothed state ($E[\boldsymbol{x}_t|\mathcal{Y}_T]$), where $\mathcal{Y}_t$ is the information provided by observations up to time t. The estimation procedure is implemented in the open source software CTSM[1] (Kristensen and Madsen (2003); Kristensen et al. (2004a)).

2.2 Transformation of the state-space

As noted the diffusion matrix should be independent of the state in order to allow the filtering equation to be simple enough to allow efficient and numerically stable solutions. Transformations of SDEs is an application of Itô's Lemma (Øksendal (2003)), the special case where the transformed system has state independent diffusion is often referred to as the Lamperti transform (Iacus (2008); Luschgy and Pagés (2006)). For one dimensional processes this is well-known (Baadsgaard et al. (1997); Iacus (2008)) and the transformation is only limited by the ability to find an explicit expression for the inverse transformation (Iacus (2008)).

Unfortunately the generality of the Lamperti transform is restricted to one dimensional diffusion (Luschgy and Pagés (2006)). It is however possible to con-

[1] Available at `www2.imm.dtu.dk/~ctsm`

struct a Lamperti type transformation for a restricted class of diffusion processes (Luschgy and Pagés (2006)) given by

$$d\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{\theta}, t)dt + \boldsymbol{\sigma}(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{\theta}, t)\boldsymbol{R}(\boldsymbol{u}_t, \boldsymbol{\theta}, t)d\boldsymbol{w}_t, \tag{6}$$

where $\boldsymbol{\sigma}(\cdot) \in \mathbb{R}^{n\times n}$ is a diagonal matrix, with diagonal elements $\sigma^{ii}(\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{\theta}, t) = \sigma^i(x_{i,t}, \boldsymbol{u}_t, \boldsymbol{\theta}, t)$ and $\boldsymbol{R}(\cdot) \in \mathbb{R}^{n\times n}$ is any matrix function (independent of $\boldsymbol{x}_t$). If z_i is chosen as

$$z_t^i = \psi^i(x_t^i, \boldsymbol{u}_t, \boldsymbol{\theta}, t) = \int \frac{d\xi}{\sigma^i(\xi, \boldsymbol{u}_t, \boldsymbol{\theta}, t)}\bigg|_{\xi=x_t^i}, \tag{7}$$

then by Itôs lemma (e.g. Øksendal (2003)), z_t^i is also an Itô process given by

$$dz_t^i = \frac{\partial}{\partial t}\psi^i(\cdot, t)dt + \frac{\partial}{\partial \xi}\psi(\xi, \cdot)\bigg|_{\xi=x_t^i} dx_t^i + \frac{1}{2}\frac{\partial^2}{\partial \xi^2}\psi^i(\xi, \cdot)\bigg|_{\xi=x_t^i} (dx_t^i)^2 \tag{8}$$

$$= \left(\psi_t(\cdot, t) + \frac{f_i(\cdot)}{\sigma^i(\cdot)}dt - \frac{1}{2}\sigma_x^i(\cdot)\sum_{j=1}^{N}\left[\boldsymbol{R}(\cdot)\right]_{i,j}^2\right)dt + \sum_{j=1}^{N}\left[\boldsymbol{R}(\cdot)\right]_{i,j} dw_j, \tag{9}$$

where the diffusion term is independent of the state. The Lamperti transformation is essentially one dimensional (Luschgy and Pagés (2006)), which is also the construction applied here. The construction (9) involves the time derivative of $\boldsymbol{\sigma}(\cdot)$. In real life application such time dependence will often be through some observed input and the time-differentiation will involve numerical differentiation of the input. It is therefore recommended that time dependence in $\boldsymbol{\sigma}(\cdot)$ is avoided if possible. Aït-Sahalia (2008) provides a more general result than (6), but (6) is simpler to apply and will suffice for our purpose.

3 Description of methodology

The procedure proposed here is an iterative procedure, where each step is repeated until an acceptable model has been achieved. The procedure is divided into 4 steps (Figure 1), the aim of step 1-3 is to identify possible model improvement, this being either model extensions or model reductions. Traditionally model extensions is implemented by formulating a hypothesis based on mechanistic knowledge and hypothesis testing by e.g. likelihood ratio testing. However, inspection of pure random walk processes adapted to data by the EKF, for different parameters can also generate data driven hypotheses, which can be tested by conventional likelihood testing.

Maximum likelihood estimation is equivalent to optimisation of the one-step predictions (to the next available observation) error. However, if the model objective is different, such as a k-step prediction or simulation, then investigating model performance with respect to this objective may lead to model extensions, potentially different from those suggested from optimising the one-step prediction.

Step 0: Data considerations

In this initial step, before model development and estimation, the main question is, if the Gaussian assumption is fulfilled. If the Gaussian assumption of observation noise is not fulfilled observations should be transformed such that the observation equation has the required form (Eq. (4)). Other considerations could be outlier detection, data aggregation, etc. These considerations should be made a priori, since changes affect the iterations in step 1-3 and especially the classical statistical hypotheses testing depend critically on the transformation of data.

Step 1: Statistical inference

This is the classical statistical step where a candidate model is formulated and statistical testing is performed by comparing likelihoods or information criteria (e.g. AIC, BIC). Possible model reductions are considered in this step, even if a model reduction seems plausible from a statistical point of view it might not be so from a modelling point of view, for instance a model reduction might lead to a model that does not fulfil basic model requirements (e.g. positive states). Additionally, some model reductions may prove unreasonable from a mechanistic understanding of the system in question, despite that statistical testing has rendered parameters equal to zero. In such cases it is preferable to maintain insignificant parameters in the model. Some of these steps are described in more details in Kristensen et al. (2004a).

If possible, model validation should also be performed by considering the autocorrelation function or generalisations like lag dependent functions (Nielsen and Madsen (2001)). These standard model validation tools are, however, not applicable for non-equidistant sampled data.

Since testing is based on one-step predictions only, rather than performance with respect to the required purpose, it might be appropriate to skip the testing

part and go directly to the validation (Step 3), when model reformulation is based on considerations with respect to the required purpose.

Step 2: Structural identification

The initial model formulation will often be simpler than the complexity of the system suggests, and the challenge is to identify possible model extensions that lead to significant improvements of the model while avoiding over-parameterisation. One way to identify potential model deficiencies is to examine the diffusion term (Kristensen et al. (2004b)), because large diffusion coefficients indicate model deficiencies in the corresponding state. Examining the observation noise may similarly pinpoint model deficiencies. Although observation noise will always be present, large observation variance suggests that the observation bear no information or alternatively that the state equations do not sufficiently describe the dynamics of the observations. Thus, the diffusion and observation noise are both expected to be positive, but large parameter estimates give hints for model improvement.

The considerations above should lead to the selection of one parameter for further analysis. This parameter is formulated as a pure random walk hidden (unobservable) state denoted by $\theta_{i,t}$,

$$d\theta_{i,t} = \sigma_{\theta_i} dw_{\theta_i}. \tag{10}$$

The model is re-estimated with the random walk diffusion (σ_{θ_i}) added to the parameter vector, and θ_i replaced by the initial state of $\theta_{i,t}$ ($\theta_{i,0}$). Ideally σ_{θ_i} should be estimated, but for complicated models this might, as we will see in Section 4.2.3, lead to small estimates on the diffusion σ_{θ_i}. Although small estimates of the diffusion term indicates that the main interactions are captured by the model, it is advisable to fix the diffusion to a moderate value that allows regular and sufficient updating of the parameter to describe parameter variations over time. The time evolution of the random walk parameter should not show systematic variations with neither time nor other states, provided that the drift term is adequately modelled.

The smoothed state or state reconstruction for the random walk parameter is calculated using the estimation procedure described above (Section 2.1), and plotted against state variables and inputs to identify possible functional relationships. Non-parametric modelling tools such as generalised additive models (GAM) (Hastie and Tibshirani (1990)) can be employed as part of this identification approach, but the significance of these relationships has to be confirmed either by testing (Step 1) or validation (Step 3).

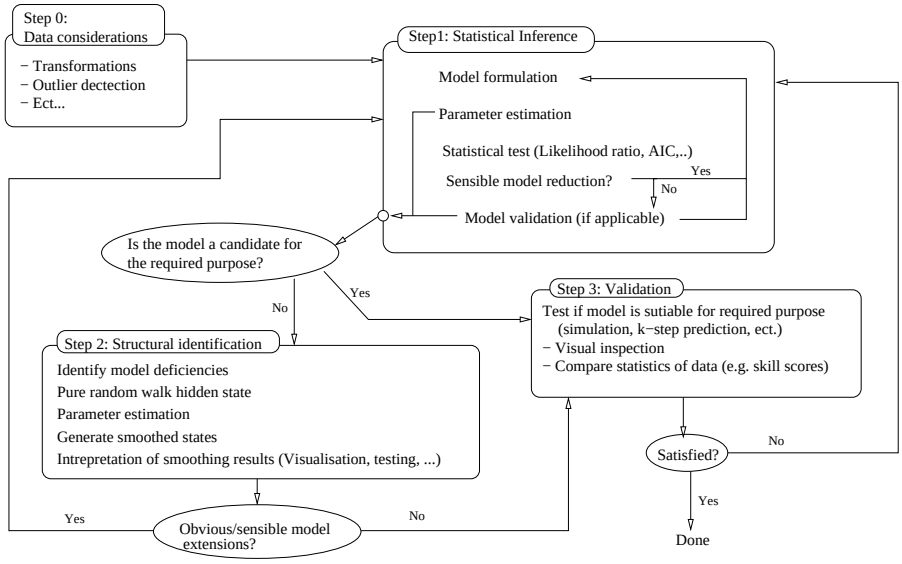

Figure 1: Conceptual diagram of model development method.

Step 3: Validation

The model developed in Step 1 and 2 should be evaluated against its objective. However, the validation methodology depends on the objective, that is, if the objective is short-term prediction then the development tools presented in Step 1 and 2 are likely to produce satisfactory results, because the likelihood estimation procedure is based on one-step predictions. On the other hand, if the model objective is long-term prediction the model developed during Step 1-2 may prove inappropriate, since model components governing the long-term predictions may not significantly affect the short-term predictions and thus may not be included in the model formulation (e.g. parameters characterising higher order moments of the stationary distribution). The general applicability of a model developed with one specific objective in mind can be assessed by various methods of cross validation, see e.g. Hastie et al. (2001) for a general discussion or Madsen (2008) for a time series oriented discussion.

In the example presented below the objective is to develop a model suitable for simulations and in this context visual inspection of the state distributions is relevant and useful for the model development. Visual inspection is, however, subjective by nature and should be combined with some kind of objective skill score. In the presented example we employ a quantile skill score and interval skill score (Gneiting and Raftery (2007)) that assigns one number to the ability

of the model to predict quantiles and confidence intervals.

4 Example: A multivariate Nitrogen-Phytoplankton model

The methodology presented above is applied to a simple two-state model for the interaction between water column nitrogen and phytoplankton, structural identification is used to generate hypothesis on primary production, these include both non-linear and multiplicative terms for nitrogen and light saturation. The model development is exemplified with water quality data from the Danish estuary, Skive Fjord, and global radiation gauged in the vicinity of Skive Fjord.

4.1 Data sources and processing

Skive Fjord is a shallow estuary located in the northern part of Denmark, which has been extensively monitored during the Danish National Aquatic Monitoring and Assessment Program (DNAMAP), and from this monitoring program measurements of total nitrogen, chlorophyll as a proxy for phytoplankton biomass and primary production sampled weekly or biweekly were used in the present study. Freshwater discharge and nitrogen input from the entire Skive Fjord watershed, calculated as combination of measured and modelled inputs, were given with a monthly resolution. Finally, global radiation with an hourly resolution was provided by the Danish Meteorological Institute (DMI).

The temporal resolution of nitrogen and freshwater inputs was increased to daily values by means of piecewise linear functions to maintain total monthly inputs. Chlorophyll data (μg chla/l) was converted into kg N/m^3 using the standard carbon to chlorophyll weight ratio of 50:1, the Redfield ratio (C:N=106:16 (molar)), and primary production (mgC/m^2) was converted to kg N/m^3 by means of the Redfield ratio and the average depth of Skive Fjord (3.2 m).

The time series for global radiation had gaps and occasionally erroneous zero values during daytime, that were also treated as missing values. Missing values were filled by linear interpolation if the sequence of missing data was short (data available within two hours from the missing observation or available at the same time of day the day before and after). After this initial gap filling, longer sequences of missing observation were filled using a general harmonic function (including a diurnal and a daily seasonal component) fitted to data. The average daily global radiation, after completing all gaps, was used as input to the

model. All data had pronounced seasonal variation, but also contained fast random variations, particularly evident for phytoplankton and primary production (Figure 2).

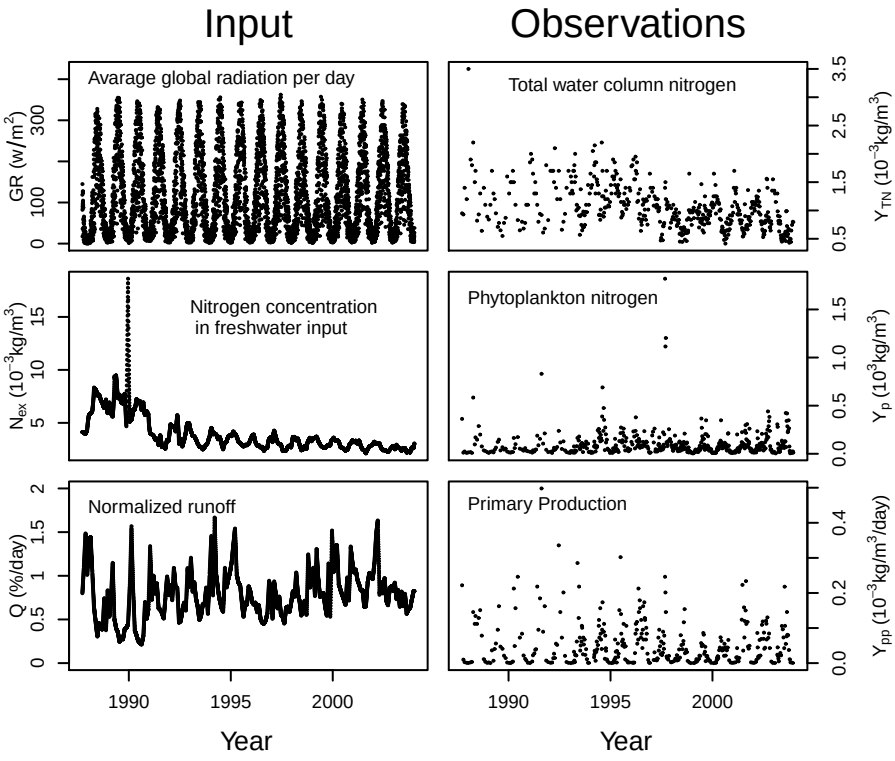

Figure 2: Data used for modelling. Input and output variables in the left panel and right panels, respectively.

4.2 The multivariate stochastic differential equation model

The conceptual setting (Figure 3) is that Skive Fjord is enriched with nitrogen discharges from the surrounding watershed (N_{ex}), whereas atmospheric deposition is relatively smaller and neglected for this model development study. To maintain the water balance of the estuary nitrogen and phytoplankton are flushed out of the system depending on the freshwater inflow (Q). The estuarine circulation will lead to additional dilution that is contained in the general loss processes, that also include denitrification and burial in the sediment (a_{wl}). a_{wl} also describes other systematic effects such as the diffusive nitrogen exchange across the sediment-water interface. Measured global radiation is a

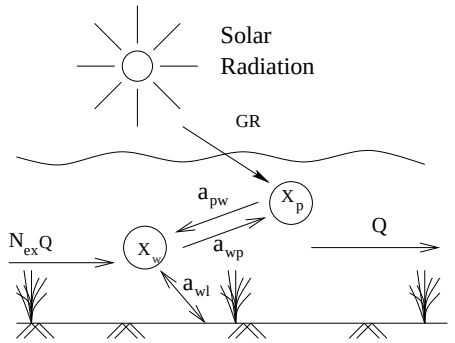

Figure 3: Conceptual diagram of the model. The state variables are X_w water column nitrogen not contained in phytoplankton, and X_p is phytoplankton nitrogen. The forcing are nitrogen input $N_{ex}Q$, nitrogen loss QX_i and the parameters are phytoplankton mortality rate a_{pw}, phytoplankton birth process a_{wp} and interactions between water column nitrogen and other compartment e.g. the sediment.

proxy for the photo-synthetic active radiation (PAR) that sustains phytoplankton growth (a_{wp}), transforming inorganic nitrogen from the water column into organic biomass. Besides the flushing described above phytoplankton loss is assumed to be mediated through the water column nitrogen (a_{pw}). Clearly, this system is a rather coarse (lumped) simplification of the many complex processes taking place, but the idea is to focus on the primary production process and lump other processes to simple first-order approximations.

In the initial step the conceptual model (Figure 3) is transformed into the simplest possible mathematical formulation encapsulating the different mass flows, thereby minimising the risk of over-parameterisation and imposing false hypotheses based on the initial model.

The minimal requirements of the mathematical formulation is that mass balance is maintained in the drift term, and that the state-space does not contain negative values. The simplest way to ensure this is by introducing noise proportional to the states, leading to the initial model formulation

$$d\begin{bmatrix} X_{w,t} \\ X_{p,t} \end{bmatrix} = \begin{bmatrix} N_{ex,t}Q_t \\ 0 \end{bmatrix} dt + \begin{bmatrix} -Q_t - a_{wp} - a_{wl} & a_{pw} \\ a_{wp} & -a_{pw} - Q_t \end{bmatrix} \begin{bmatrix} X_{w,t} \\ X_{p,t} \end{bmatrix} dt + \begin{bmatrix} \sigma_w X_{w,t} & 0 \\ 0 & \sigma_p X_{p,t} \end{bmatrix} \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} d\boldsymbol{w}_t, \tag{11}$$

where $X_{w,t}$ is the water column nitrogen not contained in phytoplankton, $X_{p,t}$ is phytoplankton nitrogen, $N_{ex,t}$ is the input of nitrogen from land, Q_t is the normalised (by the volume of Skive Fjord) freshwater input, $a_{wp}X_{w,t}$ ($a_{wp} > 0$

constant) is the primary production, $Q_t X_{i,t}$ is the flushing of nitrogen, $a_{wl} X_{w,t}$ (a_{wl} constant) is the loss/exchange of water column nitrogen through various processes besides primary production and phytoplankton mortality, $a_{pw} X_{p,t}$ ($a_{pw} > 0$ constant) is the phytoplankton mortality, $\sigma_i X_{i,t}$ ($\sigma_i > 0$) describes the system noise and r_{12} determines the noise correlation. The multiplicative noise ensures that the state-space of $\boldsymbol{X}_t = (X_{w,t}, X_{p,t})^T$ is strictly positive almost surely ($a.s.$) if $\boldsymbol{X}_0 > 0$.

Conservation of mass is not maintained in the diffusion term, and this seems reasonable because the lumped model for the loss/exchange processes in the drift term is too simple to describe in detail all the complex and interacting processes. Further, mass balance in the diffusion term would imply that random loss (gain) of phytoplankton biomass should appear in the water column, and would therefore be equivalent to only birth and death processes being stochastic. This is a quite strong assumption for a model that cannot be considered as a closed system. Additionally, mass balance in the diffusion term can to some extend be accounted for by the correlation r_{12}, and in this way it is tested by estimation.

As noted in Section 2.1, the estimation procedure does not allow state dependent diffusion, and therefore the system equation (11) is transformed according to the Lamperti transform (Eq. (7))

$$Z_{i,t} = \frac{\log(X_{i,t})}{\sigma_i} \quad \Rightarrow \quad X_{i,t} = e^{\sigma_i Z_{i,t}}, \tag{12}$$

the transformed system is given by

$$\begin{aligned} d\begin{bmatrix} Z_{w,t} \\ Z_{p,t} \end{bmatrix} = & \begin{bmatrix} \frac{1}{\sigma_w}\left(\frac{N_{ex,t}Q_t + a_{pw}X_{p,t}}{X_{w,t}} - (Q_t + a_{wl} + a_{wp})\right) & -\frac{1}{2}\sigma_w(1+r_{12}^2) \\ \frac{1}{\sigma_p}\left(a_{wp}\frac{X_{w,t}}{X_{p,t}} - (a_{pw} + Q_t)\right) & -\frac{1}{2}\sigma_p(1+r_{12}^2) \end{bmatrix} dt \\ & + \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} d\boldsymbol{w}_t, \end{aligned} \tag{13}$$

where $X_{i,t}$ is a function of $Z_{i,t}$ defined by the inverse transformation in Eq. (12).

While the state-space of the original system (11) is $[0,\infty) \times [0,\infty)$ the state-space of the transformed system is $\mathbb{R}^2$, which is tractable from an estimation point of view. The system given in (13) is observed through a set of observation equations. Under the assumption that the observations are log-normal distributed around the expectation values, these are

$$\begin{bmatrix} Y_{TN,k} \\ Y_{p,k} \\ Y_{pp,k} \end{bmatrix} = \begin{bmatrix} \epsilon_{TN,k} & 0 & 0 \\ 0 & \epsilon_{p,k} & 0 \\ 0 & 0 & \epsilon_{pp,k} \end{bmatrix} \begin{bmatrix} X_{w,k} + X_{p,k} \\ X_{p,k} \\ a_{wp} X_{w,k} \end{bmatrix}, \tag{14}$$

where $Y_{TN,k}$ is observed total nitrogen in the water column, $Y_{p,k}$ is the observed phytoplankton nitrogen, and $Y_{pp,k}$ is the observed primary production, all at time $t = t_k$. In the log-domain we obtain the observation equation

$$\begin{bmatrix} \log(Y_{TN,k}) \\ \log(Y_{p,k}) \\ \log(Y_{pp,k}) \end{bmatrix} = \begin{bmatrix} \log(X_{w,k} + X_{p,t_k}) \\ \log(X_{p,k}) \\ \log(a_{wp} X_{w,k}) \end{bmatrix} + \begin{bmatrix} e_{TN,k} \\ e_{p,k} \\ e_{pp,k} \end{bmatrix}, \tag{15}$$

where $X_{w,k}$ and $X_{p,k}$ are described by the inverse transformation Eq. (12), and $e_{i,k} \sim N(0, s_i^2)$. In order to strengthen conclusions obtained in the structural identification step we have chosen to use all available data in for estimation and the validation described Section 4.3.1 is therefore based on the same data (but not on likelihood performance).

4.2.1 Model 1: The linear model

The parameters of the linear model Eqs. (13) and (15) were estimated. All parameters, except a_{pw}, display good estimation statistics (Table 1). Although a_{pw} is not statistically significant at this step of the modelling procedure, it is not advisable to remove it, since this would lead to a biologically meaningless model with a zero death rate. The estimated correlation coefficient is negative, implying that part of the randomness introduced by the diffusion is affecting the difference between primary production and mortality.

The model does not implicitly contain any other seasonal elements than the nitrogen input (N_{ex}) to describe the strong seasonality of the data displayed in Figure 2. The problem of how to identify possible improvements of the model Eqs. (11) and (14) in an objective way is the main theme of this paper. While the parameterisation of the diffusion term is not extremely important for the estimation problem as such, because the process is kept in place by the EKF, the diffusion term is a main driver for the distributional properties of the process and will be addressed in Section 4.4.1. The diffusion estimates (Table 1) highlights that the phytoplankton diffusion (σ_p) is approximately 10 times larger than the water column nitrogen diffusion (σ_w). Further, the observation noise of primary production (s_{pp}^2) is very large, which makes it reasonable to select the primary production parameter (a_{wp}) for a more thorough examination. This is also a plausible hypothesis from a biological point-of-view since primary production is traditionally modelled as a function of PAR and phytoplankton biomass as well as available nitrogen. For notational convenience, we will represent the primary production process by

$$a_{wp,t}^i = a_{wp}^0 X_{w,t} f_i(\boldsymbol{X}_t, \boldsymbol{u}_t), \tag{16}$$

Table 1: Estimation results for Model 1-6, bold face number refer to significant (on a 5% level) parameters, while number in parenthesis refer to standard deviation of parameter estimates.

	Unit	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
				Drift parameters			
a^0_{wp}	[a] d^{-1}	**0.016**	**0.159**	**0.361**	**0.376**	**0.362**	**0.405**
		(0.002)	(0.006)	(0.020)	(0.021)	(0.020)	(0.024)
a_{wl}	d^{-1}	**0.007**	**0.018**	**0.018**	**0.018**	**0.02**	**0.024**
		(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
a_{pw}	d^{-1}	0.053	**0.124**	**0.319**	**0.310**	**0.322**	**0.379**
		(0.034)	(0.007)	(0.019)	(0.020)	(0.018)	(0.021)
k_w	$\frac{g}{m^{-3}}$			**0.067**	**0.106**	0.025	0.006
				(0.017)	(0.020)	(0.025)	(0.020)
k_{gr}	$\frac{W}{m^{-3}}$				**0.990**	**9.283**	**18.068**
					(0.244)	(3.247)	(3.360)
a_{p0}	$\frac{g}{m^3 \cdot d}$					**0.002**	**0.004**
						(0.001)	(0.001)
				Diffusion parameters			
σ_w	[b] $\frac{g}{m^3\sqrt{d}}$	**0.061**	**0.055**	**0.063**	**0.063**	**0.061**	**0.062**
		(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
σ_p	[b] $\frac{g}{m^3\sqrt{d}}$	**0.625**	**0.252**	**0.152**	**0.145**	**0.197**	**0.065**
		(0.085)	(0.015)	(0.008)	(0.009)	(0.017)	(0.012)
r_{12}		**−0.164**	**−0.135**	−0.057	−0.046	0.014	0.009
		(0.046)	(0.034)	(0.035)	(0.039)	(0.039)	(0.042)
γ_w							**0.662**
							(0.128)
γ_p							**0.546**
							(0.062)
				Variance of observation noise			
s^2_{TN}	$\frac{g^2}{m^6}$	**0.013**	**0.029**	**0.013**	**0.013**	**0.012**	**0.011**
		(0.002)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)
s^2_p	$\frac{g^2}{m^6}$	**0.304**	**0.141**	**0.172**	**0.185**	**0.198**	**0.205**
		(0.148)	(0.047)	(0.018)	(0.036)	(0.022)	(0.031)
s^2_{pp}	$\frac{g^2}{m^6 d^2}$	**3.546**	**2.193**	**1.174**	**1.126**	**0.851**	**0.715**
		(0.285)	(0.179)	(0.095)	(0.098)	(0.102)	(0.075)

[a] Unit does not apply to Model 2, where the unit is $\frac{m^3}{g \cdot d}$.

[b] Units does not apply to Model 6, where the unit of σ_i is $\left(\frac{g}{m^3\sqrt{d}}\right)^{-\gamma_i}$.

Table 2: Likelihood table or Model 1-6, column 1-4 refer to the model set up, column 5 report the log-likelihood, column 6 report the total number of degrees of freedom (including the initial state), while column 6-7 report AIC and BIC for all models and the last column reports the likelihood ratio test when applicable.

	f	a_{p0}	γ_i	$\log(L)$	DF	AIC	BIC	$P(x > -2Dl)$
Model 1	$f_1 = 1$	0	1	-1446	11	2914	2971	
Model 2	$f_2 = f_1 X_{p,t}$	0	1	-1374	11	2770	2828	
Model 3	$f_3 = \frac{f_2}{k_w + X_{w,t}}$	0	1	-1079	12	2182	2245	
Model 4	$f_4 = \frac{f_3 GR_t}{k_{gr} + GR_t}$	0	1	-1067	13	2159	2227	
Model 5	$f_5 = f_4$	free	1	-1007	14	2042	2115	0.0000
Model 6	$f_6 = f_4$	free	free	-986	16	2004	2088	0.0000

where f_i is the functional expression to be identified (with $f_1(\boldsymbol{X}_t, \boldsymbol{u}_t) = 1$ in the linear model). In each step, i, a^0_{wp} is replaced with $a^0_{wp,t}$ to identify a candidate model, $i + 1$ (replacing $f_i(\cdot)$ with $f_{i+1}(\cdot)$).

The primary production parameter ($a^0_{wp,t}$) is now modelled as a random walk process that will adapt to the data through the EKF. To ensure $a^0_{wp,t} > 0$, $\forall t$ the random walk is introduced in the log domain, resulting in the following equation for the random walk parameter

$$d\log(a^0_{wp,t}) = \sigma_{a_{wp}} dw_{a_{wp},t}. \tag{17}$$

The parameters of this modified model description and the smoothened state $E(\boldsymbol{X}_t|\mathcal{Y}_T)$ (the mean of posterior distribution) from the EKF are estimated. A clear seasonal variation in $\hat{a}^0_{wp,t|T}$ (Figure 4A) as well as evident correlations with phytoplankton nitrogen ($X_{p,t}$), water-column nitrogen ($X_{w,t}$), and global radiation GR_t (Figure 5A-C) emerges, however most pronounced with phytoplankton. This is also confirmed by comparing AIC for linear models of the hypothesised relationships

$$\hat{a}^0_{wp,t|T} = \alpha_{gr,0} + \alpha_{gr,1} Gr_t + \epsilon_{Gr,t} \qquad AIC = -22928 \tag{18}$$

$$\hat{a}^0_{wp,t|T} = \alpha_{p,0} + \alpha_{p,1} \hat{X}_{p,t|T} + \epsilon_{p,t} \qquad AIC = -28410 \tag{19}$$

$$\hat{a}^0_{wp,t|T} = \alpha_{w,0} + \alpha_{w,1} \hat{X}_{w,t|T} + \epsilon_{w,t} \qquad AIC = -23998. \tag{20}$$

4.2.2 Model 2: Including phytoplankton

Based on the statistics in Eqs. (18)-(20) and Figure 5A-C phytoplankton nitrogen is included in the production process and hence $f_2(\cdot)$ (in Eq. (16)) is

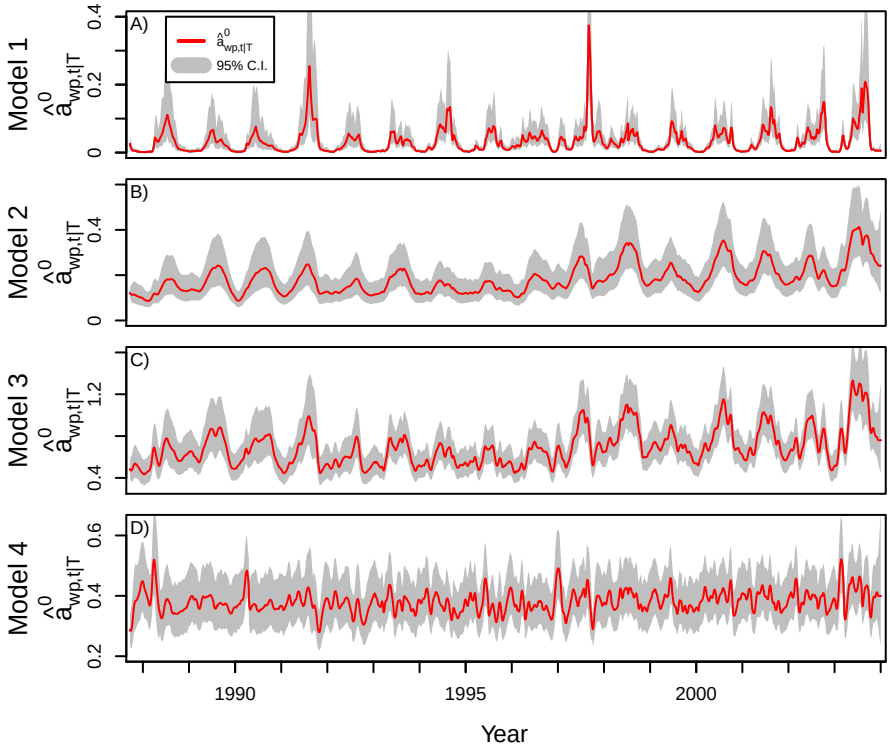

Figure 4: Smoothened state of the random walk phytoplankton growth parameter ($\hat{a}^0_{wp,t|T}$) in Model 1-4 as function of time. 95% confidence interval (grey area) is base on a Gaussian assumption of $\log(\hat{a}^0_{wp,t|T})$. Red lines represent the median of $\hat{a}^0_{wp,t|T}$.

modelled as

$$f_2(\boldsymbol{X}_{t_k}, \boldsymbol{u}_t) = X_{p,t}. \tag{21}$$

Model 1 is not a proper subset of Model 2 implying that likelihood ratio testing is not valid. We therefore base the model evaluation on AIC and BIC, improvements in both criteria in the order of 140 (Table 2) are seen. All parameters in this model formulation, including the death rate, are well determined (Table 1), though the phytoplankton diffusion (σ_p) and primary production observation noise (s^2_{pp}) are still large compared to the diffusion of X_w, despite that both decreased.

To address these large errors the primary production parameter is again modelled as a random walk and plotted as a function of the potential explanatory

variables and time (Figures 5D-F and 4B). A clear seasonal variation (Figure 4B) still remains as well as evident correlation between $\hat{a}^0_{wp,t|T}$ and the state variables and global radiation (Figure 5D-F). It is also seen that linear relations would be a poor fit (Figure 5D-F), and the following hypotheses are therefore considered

$$H_1 : a_{wp} = \frac{Gr_t}{k_{gr} + Gr_t} \tag{22}$$

$$H_2 : a_{wp} = \frac{X_{p,t}}{k_p + X_{p,t}} \tag{23}$$

$$H_3 : a_{wp} = \frac{1}{k_w + X_{w,t}}. \tag{24}$$

H_1 is based on Figure 5F and the well-known fact that light saturation occurs for primary production, although the parametric description here is simpler than the usual parameterisation of light saturation (e.g. Fasham et al. (1990)). H_2 is based on empirical evidence only (Figure 5D), while H_3 is based on Figure 5E and Michaelis-Menten kinetics for nitrogen. These hypotheses are transformed into linear relations and the best relationship based on AIC is chosen

$$\frac{1}{a_{wp,t}} = \alpha_{gr,1}\frac{1}{Gr_t} + \alpha_{gr,2}Gr_t + \epsilon_{Gr,t} \qquad AIC = 34812 \tag{25}$$

$$\frac{1}{a_{wp,t}} = \alpha_{p,1}\frac{1}{X_{p,t}} + \alpha_{p,1}X_{p,t} + \epsilon_{p,t} \qquad AIC = 30108 \tag{26}$$

$$\frac{1}{a_{wp,t}} = \alpha_{w,0} + \alpha_{w,1}X_{w,t} + \epsilon_{w,t} \qquad AIC = 15040. \tag{27}$$

The lowest AIC was the formulation with water column nitrogen Eq. (27), which is also most apparent from the less scatter in Figure 5E.

4.2.3 Model 3: Including nitrogen saturation

Based on the statistics in Eqs. (25)-(26) and Figure 5D-F, water column nitrogen is included in the model, and the primary production process is reformulated as

$$f_3(\boldsymbol{X}_{t_k}, \boldsymbol{u}_t) = \frac{X_{p,t}}{k_w + X_{w,t}}. \tag{28}$$

This model, now including nitrogen saturation of primary production, was first fitted without random walk parameters (Table 1). All parameters, except the correlation parameter, were significant, even though the statistics suggest to remove the correlation parameter, it is maintained in the model and the model reduction step is postponed to the final model. Since Model 2 is not a proper

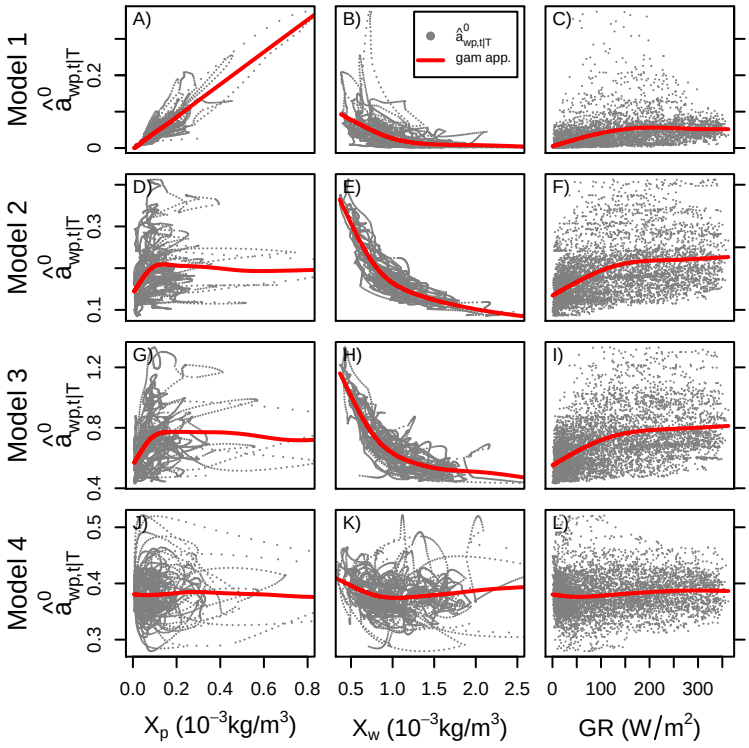

Figure 5: Smoothened state of the random walk phytoplankton growth parameter ($\hat{a}^0_{wp,t|T}$) in Model 1-4 as a function of smoothen phytoplankton level (first column), smoothen water column nitrogen (second column), observed global radiation (third column). Grey dots represent point estimation of $\hat{a}^0_{wp,t|T}$, while red lines represent GAM (smoothing splines) fit to the points.

subset of Model 3 model evaluation is again based on AIC and BIC, improvements was about 600 for both criteria (Table 2). It should also be stressed that the diffusion of X_p and the observation noise of primary production (s^2_{pp}) both decreased.

The estimated random walk diffusion for the primary production parameter ($a^0_{wp,t}$) almost disappeared ($2.4 \cdot 10^{-4}$) after this model change, and there was no seasonal variation in the random walk, despite an anticipated seasonal pattern yet uncovered in the model formulation. To address this artefact from the estimation procedure and allow regular updating of $a^0_{pw|t}$, the diffusion for the random walk parameter was fixed to 0.05, which is comparable to the diffusion of the water column diffusion parameter. Following this modification the strongest

relationship to the random walk parameter is still with water column nitrogen (Figure 5H), however, it is also evident that the random walk parameter was related to global radiation (Figure 5I). As the intention is to build a model that is well suited for simulation, it is necessary to include a seasonal input, but also cross correlation between X_w and GR might influence the results seen in Figure 5G-I. Therefore global radiation is included, acknowledging that the relationship between $a^0_{pw|t}$ and GR is clearly not linear, the simple light saturation function given in Eq. (22) is chosen.

4.2.4 Model 4: Including global radiation

Based on the reasoning above $f_4(\cdot)$ is chosen as

$$f_4(\boldsymbol{X}_t, \boldsymbol{u}_t) = \frac{Gr_t X_{p,t}}{(k_{gr} + Gr_t)(k_w + X_{w,t})}. \tag{29}$$

Again all parameters except for the correlation coefficient are clearly significant (Table 1). The phytoplankton diffusion (σ_p) and primary production observation noise (s^2_{pp}) did not decrease (Table 1), but the improvements of AIC and BIC is about 20 (Table 2).

The primary production parameter is again modelled as a random walk with fixed diffusion parameter equal 0.05. There is no strong correlation between the states and the input (Figure 5J-L), and the distinctive seasonal pattern observed in Figure 4A-C has disappeared (Figure 4D). As the developed model is a potential candidate for simulation studies, we continue to Step 3 in the model development procedure to validate if the model is suitable for simulation.

4.3 Simulations

The modelling so far has focused on formulation hypothesis based on pure random walk primary production parameter, likelihood testing and information criteria. Likelihood testing is equivalent to optimisation of the ability to predict the observation at time t_{k+1} given the information up to time t_k. The time between water quality samples vary from 4 to 94 days with an average sampling time ranging from 11.5 to 16 days for the different water quality variables. For simulation studies the objective of the model is to predict perhaps one or several years ahead. Further the aim of a simulation is not only to predict one value (like the mean) of future states, but to predict the distribution of the future states. This imply that we need to get both the drift and the diffusion term right. The simulation studies in the following is based on the Euler scheme (Kloeden and

Platen (1999)) applied to the transformed process with $\Delta t = \frac{1}{96}d$. The time step Δt is decreased until the simulation plots (like Figure 6) do not change.

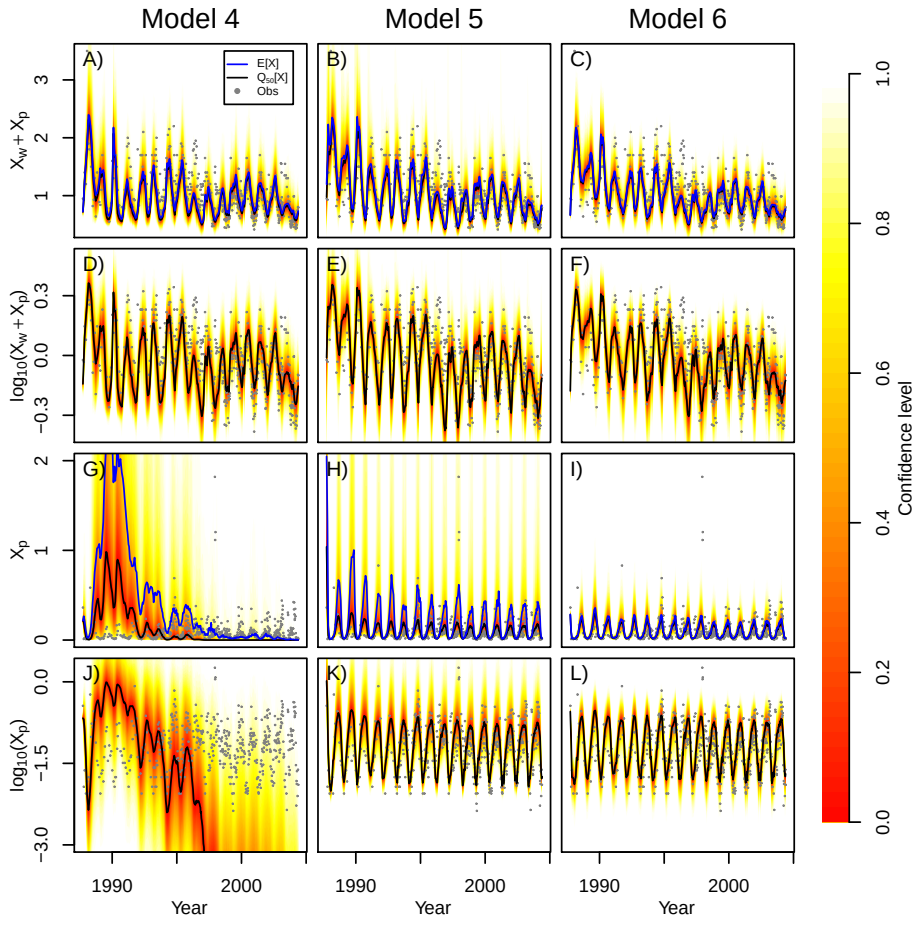

Figure 6: Simulation for the time span of observations with Model 4-6, colour code refer to confidence intervals around the median of the distribution, while gray dots are the measurements.

4.3.1 Simulation of Model 4

Model 4 was simulated over the entire span of the dataset, with the initial state drawn from a Gaussian distribution around the smoothened state, having a mean and variance equal to that of the smoothen state (in the transformed do-

main). The simulated water column nitrogen seems to be well captured (Figure 6A and D), while phytoplankton (Figure 6G and J) and primary production (not shown, but similar to the phytoplankton plot) perform poorly, with the simulated distribution drifting away from the observations. To address this problem, we reconsider the developed phytoplankton equation

$$\begin{aligned} dX_{p,t} = &\left(a_{wp} \frac{X_{w,t} G r_t}{(k_w + X_{w,t})(k_{gr} + G r_t)} - Q_t - a_{pw} \right) X_{p,t} dt \\ &+ \sigma_p X_{p,t} (r_{12} dw_{1,t} + dw_{2,t}). \end{aligned} \tag{30}$$

The stationary distribution (considering X_w, Q_t and GR_t as constants) for this equation is either 0 or ∞ depending on the factor in front of $X_{p,t}dt$. Such a behaviour is clearly not desirable and is the main course of the drift seen in Figure 6G and J. In order to solve this problem we add a constant ($a_{p0} > 0$) to the equation to get

$$\begin{aligned} dX_{p,t} = &a_{p0} dt + \left(a_{wp} \frac{X_{w,t} G r_t}{(k_w + X_{w,t})(k_{gr} + G r_t)} - Q_t - a_{pw} \right) X_{p,t} dt \\ &+ \sigma_p X_{p,t} (r_{12} dw_{1,t} + dw_{2,t}). \end{aligned} \tag{31}$$

The argumentation for the constant a_{p0} is clearly mathematical convenience, but we can think of a_{p0} as an inoculum. Also if the constant is small it will not influence (local) phytoplankton growth greatly.

4.4 Model 5: Including a constant inoculum

The model with a constant inoculum factor is also well determined (Table 1) and the likelihood improved by 59.9 on 1 degree of freedom (Model 4 is a proper subset of Model 5), which is significant with $p << 0.0001$. To evaluate if a_{p0} is small, imagine that no water column nitrogen is present and that $Q_t = 0$, then the mean value (of the stationary distribution) of the phytoplankton would be $\frac{a_{p0}}{a_{pw}} = 6.8 \cdot 10^{-3} \frac{g}{m^3}$ (Forman and Sørensen (2008); Iacus (2008)), which is low compared to the observed values (about 0.7% of the observed phytoplankton nitrogen is below this value). The nitrogen saturation k_w constant is not significant in t-test (Table 1), and it is low compared to the observed values of water column nitrogen, it is however comparable to the saturation constant used by Fasham et al. (1990) (0.007 gNm^{-3}) and the reported saturation constant by Fisher et al. (1992) (0.028 gNm^{-3}). Furthermore if k_w is removed the model would be able to produce undesirable negative values of water column nitrogen, and k_w is therefore kept in the model.

An important test of Models 5 is its behaviour in a simulation study. The annual variations of the phytoplankton state is captured much better (Figure 6H and

K), however the model does predict very large mean values for phytoplankton (Figure 6H) and primary production (not shown, but similar to the phytoplankton plot), and the confidence intervals are also large, with values exceeding the largest observations.

4.4.1 Model 6: Analysis of diffusion

These wide confidence intervals obtained with Model 5 could be due to the diffusion scaling too fast with the state. In order to analyse this question, we replace the diffusion term, in Eq. (11), with

$$\begin{bmatrix} \sigma_w X_{w,t}^{\gamma_w} & 0 \\ 0 & \sigma_p X_{p,t}^{\gamma_p} \end{bmatrix} \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}. \tag{32}$$

The coefficients γ_w and γ_p are commonly chosen as either 1 (Øksendal (2003)) or 0.5 (Klebaner (2005)) in biological models. For γ_i equal to 0.5 the linear one-dimensional process is known as Feller diffusion in biology (Klebaner (2005)) or CIR model in finance (Iacus (2008)). It has a positive probability of reaching zero if the loading parameter is small compared to the diffusion parameter (Iacus (2008)), while for γ_i larger than 1 the system does not fulfil the linear growth condition, and existence and uniqueness is not guaranteed (Øksendal (2003)).

The Lamperti transform presented in Eq. (12) needs to be reformulated as

$$Z_{i,t} = \frac{X_{i,t}^{1-\gamma_i}}{\sigma_i(1-\gamma_i)} \quad \Rightarrow \quad X_{i,t} = [\sigma_i(1-\gamma_i)Z_{i,t}]^{\frac{1}{1-\gamma_i}}. \tag{33}$$

For $\gamma_i \in (0,1)$ the state-space of $Z_{i,t}$ is equal to the state-space of $X_{i,t}$ ($[0,\infty)$). The transformed system is given by

$$\begin{aligned} dZ_{w,t} =& \frac{X_{w,t}^{-\gamma_w}}{\sigma_w}\left(N_{ex,t}Q_t - \left(Q_t + a_{wl} + a_{wp}^0 f_4(\boldsymbol{X}_t, \boldsymbol{u}_t)\right) X_{w,t} + a_{pw}X_{p,t}\right) dt \\ & - \frac{1}{2}\sigma_w\gamma_w X_{w,t}^{\gamma_w - 1}(1+r_{12}^2)dt + dw_1 + r_{12}dw_2 \end{aligned} \tag{34}$$

$$\begin{aligned} dZ_{p,t} =& \frac{X_{p,t}^{-\gamma_p}}{\sigma_p}\left(a_{p0} + a_{wp}^0 f_4(\boldsymbol{X}_t, \boldsymbol{u}_t)X_{w,t} - (a_{pw} + Q_t)X_{p,t}\right) dt \\ & - \frac{1}{2}\sigma_p\gamma_p X_{p,t}^{\gamma_p - 1}(1+r_{12}^2)dt + r_{12}dw_1 + dw_2. \end{aligned} \tag{35}$$

Rearranging and using the short hand notation, $\tilde{\theta}_{w,0} = (N_{ex,t}Q_t + a_{pw}X_{p,t})/\sigma_w$, $\tilde{\theta}_{w,1} = (Q_t + a_{wl} + a_{wp}^0 f_4(\boldsymbol{X}_t, \boldsymbol{u}_t))/\sigma_w$, and $\tilde{\theta}_{w,2} = \frac{1}{2}\sigma_w\gamma_w(1+r_{12}^2)$ gives the following SDE for the water column nitrogen

$$dZ_{w,t} = X_{w,t}^{-\gamma_w}\left(\tilde{\theta}_{w,0} - \tilde{\theta}_{w,1}X_{w,t} - \tilde{\theta}_{w,2}X_{w,t}^{2\gamma_w - 1}\right) dt + dw_1 + r_{12}dw_2. \tag{36}$$

Now consider the limit where $X_{w,t} \to 0$, for $\gamma_w \neq \frac{1}{2}$ we get

$$\lim_{X_w \to 0} dZ_{w,t} = \lim_{X_w \to 0} \begin{cases} -\tilde{\theta}_{w,2} X_{w,t}^{\gamma_w - 1} dt & = -\infty dt \quad a.s. \quad \gamma_w < \frac{1}{2} \\ \tilde{\theta}_{w,0} X_{w,t}^{-\gamma_w} dt & = +\infty dt \quad a.s. \quad \gamma_w > \frac{1}{2}. \end{cases} \tag{37}$$

For $\gamma_w = \frac{1}{2}$ the limit splits into three cases

$$\lim_{X_w \to 0} dZ_{w,t} = \lim_{X_w \to 0} \begin{cases} (\tilde{\theta}_{w,0} - \tilde{\theta}_{w,2}) X_{w,t}^{-\frac{1}{2}} dt & = -\infty dt \quad a.s. \quad \tilde{\theta}_{w,0} < \tilde{\theta}_{w,2} \\ dw_1 + r_{12} dw_2 & \qquad\qquad\quad \tilde{\theta}_{w,0} = \tilde{\theta}_{w,2} \\ (\tilde{\theta}_{w,0} - \tilde{\theta}_{w,2}) X_{w,t}^{-\frac{1}{2}} dt & = +\infty dt \quad a.s. \quad \tilde{\theta}_{w,0} > \tilde{\theta}_{w,2}. \end{cases} \tag{38}$$

These considerations show that the transformed system is not consistent with the topology of the state-space when $\gamma_w < \frac{1}{2}$, and for $\gamma_w = \frac{1}{2}$ the transformed system is only consistent with the topology of the state-space for a restricted set of parameter values ($\tilde{\theta}_{w,0} > \tilde{\theta}_{w,2}$). The reasons for these inconsistencies is that the transformation is only valid inside the state-space and not on the boundary. The parameter set should reflect that $P(X_{w,t} = 0) = 0$. Clearly the simplest way to ensure this is by restricting γ_w to the interval $(0.5, 1)$. Similar arguments apply to the phytoplankton equation, and γ_p is therefore also restricted to the interval $(0.5, 1)$.

The parameters were estimated with $\gamma_i \in (0.5, 1)$. Most parameters are well determined (Table 1) (with the exception of the correlation coefficient and nitrogen saturation), and further the diffusion coefficient of X_p and X_w have comparable sizes. The simulation study (Figure 6 third column) shows that the confidence intervals for phytoplankton nitrogen has narrowed (Figure 6I and L) and that phytoplankton mean is now close to the median of the distribution. The consequence is that more extreme values are not included in the distribution of the simulations.

4.5 Quantification of simulation performance

The purpose of simulation studies like the ones presented in Figure 6 is to predict the distribution of the future state of the system. One way of quantifying this analysis is to compare simulation quantiles with observed data. The visual inspection of the simulation results (Figure 6) is clearly relevant and led to the introduction of a_{p0} and γ_i. However, to quantify the model skills we need to represent the performance by a single number, in the same way as the likelihood values represent the overall quality of short-term predictions.

A good model candidate should 1) be reliable meaning the quantiles of the estimated distribution should hold the right proportion of the data (often referred

to as reliability) and 2) have narrow confidence regions (often referred to as sharpness (Gneiting and Raftery (2007)). In addition, the ability to adapt to different uncertainty regimes is sometimes considered (often referred to as resolution). Though misleading results may emerge if sharpness and resolution are considered only and reliability is not taken into account (Møller et al. (2008); Pinson et al. (2007)). A proper skill score for quantiles should therefore combine these objectives into one single number. One such skill score is (Gneiting and Raftery (2007))

$$S(r_1, .., r_k; x) = \sum_{i=1}^{k} (\alpha_i s_i(r_i) + (s_i(x) - s_i(r_i))\mathbb{I}\{x \leq r_i\}) + h(x), \qquad (39)$$

where x is the observation, r_i is the quantile predicted by the simulation, s_i is non-decreasing and h is arbitrary. Here we choose $s_i(x) = x$ and $h(x) = -\sum_i \alpha_i x$, and in this case Eq. (39) is the sum of loss functions for quantile regression as defined by Koenker and Basset (1978). Model 6 perform consistently better than Model 4 and 5 when comparing the skill score for all observations (Table 3). Based on these statistics Model 6 is the better choice of model.

To evaluate individual confidence intervals the (negatively oriented) interval score (Gneiting and Raftery (2007)) is calculated by

$$S_\alpha^{int}(l, u; x) = (u - l) + \frac{2}{\alpha}(l - x)\mathbb{I}\{x < l\} + \frac{2}{\alpha}(x - u)\mathbb{I}\{x > u\}, \qquad (40)$$

where $[l, u]$ is the confidence interval and α is the nominal coverage. As noted by Gneiting and Raftery (2007) this is intuitively appealing, since sharpness is formulated directly $(u-l)$ and values outside the interval are penalised. Model 6 performs consistently better for total nitrogen and primary production (Figure 7), Model 5 performs slightly better for high confidence levels of phytoplankton (above 0.6) while Model 6 performs better for confidence levels below 0.6. The combined conclusion is therefore that Model 6 is the preferred candidate for the final model.

The better performance of Model 5 compared to Model 6, in terms of nominal coverage of phytoplankton nitrogen (Figure 7), reflects that extreme values are not well described in Model 6. This is not surprising as neither Model 6 nor any of the other models contain any mechanisms to specifically capture these extreme events. To capture these extremes, information on e.g. oxygen depletion would be needed, but oxygen depletion events are governed by local weather conditions such as wind, which cannot be predicted long time ahead. It is therefore not realistic to predict extremes with a simulation model that runs for several of years.

Table 3: Quantile skill scores and quantile skill score differences for the simulation models. Skill scores are calculated in the log-domain, and index refer to observations.

	S_{TN}	DS_{TN}	S_p	DS_p	S_{pp}	DS_{pp}
Model 4	-8.96		-214.32		-187.96	
Model 5	-8.98	-0.02	-24.00	190.32	-30.34	157.62
Model 6	-8.39	0.60	-23.65	0.35	-28.29	2.05

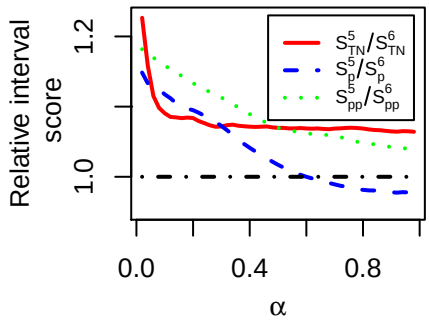

Figure 7: Relative interval quantile skill score for simulation models 5 and 6; α refer to nominal coverage. Values above 1 indicate better performance of Model 6 while values below 1 indicate better performance of Model 5.

4.6 Model reductions

Model 6 is the candidate for the final model formulation, although the results in Table 1 suggest two potential model reductions. Nitrogen half saturation (k_w) is not significant, however, as discussed above, it is not reasonable to remove this parameter since it will lead to a positive probability of reaching negative values for water column nitrogen. It should be noted that an attempt to estimate parameters in a Model 6 with $k_w = 0$ failed, because the optimiser tested values that led to negative values of water column nitrogen.

The correlation coefficient (r_{12}) is also not significant, and consequently this parameter is removed and the model re-estimated, yielding a log-likelihood decrease of 0.023, which corresponds to a p-value of 0.83. Furthermore, the relative (to the standard deviation) changes of the parameters are all less than 0.1. Thus

the final model consists of the syatem equation

$$d\begin{bmatrix} X_{w,t} \\ X_{p,t} \end{bmatrix} = \begin{bmatrix} N_{ex,t}Q_t - (Q_t + a_{wl})X_{w,t} - \frac{a^0_{wp}X_{w,t}X_{p,t}GR_t}{(k_w+X_{w,t})(k_{gr}+GR_t)} + a_{pw}X_{p,t} \\ a_{p0} + \frac{a^0_{wp}X_{w,t}X_{p,t}GR_t}{(k_w+X_{w,t})(k_{gr}+GR_t)} - (a_{pw} + Q_t)X_{p,t} \end{bmatrix} dt + \begin{bmatrix} \sigma_w X^{\gamma_w}_{w,t} & 0 \\ 0 & \sigma_p X^{\gamma_p}_{p,t} \end{bmatrix} d\boldsymbol{w}_t, \quad (41)$$

and the observation equation

$$\begin{bmatrix} \log(Y_{TN,k}) \\ \log(Y_{p,k}) \\ \log(Y_{pp,k}) \end{bmatrix} = \begin{bmatrix} \log(X_{w,k} + X_{p,t_k}) \\ \log(X_{p,k}) \\ \log(a_{wp}X_{w,k}) \end{bmatrix} + \begin{bmatrix} e_{TN,k} \\ e_{p,k} \\ e_{pp,k} \end{bmatrix}. \quad (42)$$

4.7 Discussion of the biological model

The model development presented here is based primarily on mathematical and statistical reasoning, while the biological/physical reasoning is mostly used in the initial model formulation. The model development based on visual inspection of the (smooth) path of the primary production parameter is, however, consistent with existing knowledge from N-P models. Identified model extensions based on the random walk hidden states are all significant, as are the extensions based on considerations of the simulated or stationary distribution.

Clearly, the presented model is very simple and represents a coarse simplification of the complex ecosystem, and a higher dimensional state-space would be needed to give a realistic description owing to all the known processes. The simplicity is a deliberate choice, since the focus is placed on demonstrating the model development and the significant role of the diffusion term. The structural identification and the considerations of the simulated distributions are, however, not limited to simple systems and the identification methodology can principally be applied directly to models with a higher number of state variables as well. Further, the probabilistic formulation of state transitions also lump weak interactions that may otherwise be formulated in a complicated ODE model, often with a large degree of uncertainty. The SDE setting provides a direct quantification of the uncertainty, which cannot be estimated directly from ODE models. The model development does not include explicit knowledge about the parameters, and prior knowledge is only included implicitly in the hypothesis formulations.

The key importance of the diffusion term is illustrated by the large improvements in the likelihood criteria, when introducing the exponents γ_i, stressing the

importance of the specific parametric formulation of diffusion term. One of the main results of the model development is the large reductions of the confidence band width of the simulated densities obtained by introducing γ_i. Actually, a hypothesis of $\gamma_i = \frac{1}{2}$ cannot be rejected using unconditional t-tests for each of the exponents, and the results therefore support the hypothesis that the diffusion is of the Feller type. It is, however, argued that estimating γ_i in the open interval $\left(\frac{1}{2}, 1\right)$ is robust from a estimation point of view, but just as important is that strictly positive states included in the drift term are maintained in the stochastic formulation when $\gamma_i > \frac{1}{2}$.

The model focuses mainly on primary production, whereas loss processes are lumped together. The partitioning of the loss term in the model could be carried out following the same methodology as presented for the production process, provided that explanatory information or observations needed to describe the different terms (e.g. zooplankton and filter feeder biomass) is available. This will probably result in a more variable set of loss functions as opposed to the constant loss rate (a_{wl}) used in the present context. Another issue is that the model does not distinguish between labile and non-labile nitrogen for phytoplankton growth, and a better partitioning of the nitrogen pool may lead to further model improvements. All these potential improvements imply additional states in the system, which makes both estimation and inferences more complicated, and therefore render such model extensions less appropriate for introducing the methodology. More importantly, a more detailed process description requires additional information that is not available in standard monitoring programmes.

The light half-saturation parameter is well determined by the procedure. Comparing the half-saturation parameter (Figure 8) with the range of data shows that there are observations on both side of the constant. In contrast, the half-saturation parameter for water column nitrogen is far below any observed value, implying that phytoplankton growth is not severely limited by nitrogen. Skive Fjord is a eutrophic ecosystem with large land-based inputs of nitrogen, and ambient concentrations of dissolved inorganic nitrogen are mostly above the levels reported to limit phytoplankton growth in experiments (e.g. Fisher et al. (1992)).

5 Conclusion

The methodology presented in this study is based on likelihood estimation for identifying probabilistic models of physical/biological phenomena, formulated as a system of Itô stochastic differential equations partially observed in discrete

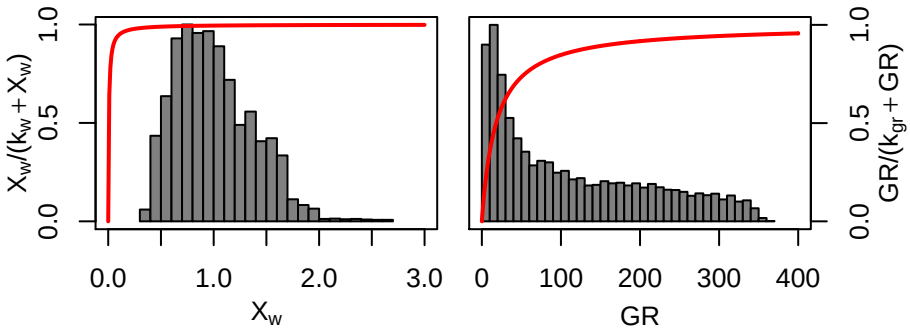

Figure 8: Nonlinear multiplicative effects for nitrogen and light saturation. Red lines show the estimated relationships, and the histogram show the distributions of water column nitrogen (left) and observed global radiation (right).

time through a set of observation equations. By formulating parameters of the stochastic differential equation model as pure random walk hidden (unobserved) states, it is demonstrated how embedded structural information can be extracted by analysis of the smoothen state estimates. The selection of which parameter should be analysed to improve the model is based on considerations about the diffusion matrix *and* the observation covariance matrix. Large values of either or both of these terms for a particular state propose model deficiencies in the corresponding state and therefore possible model improvements. The diffusion of the random walk hidden state parameter should preferably be estimated to describe the dynamics of the parameter. It is demonstrated that even when the random walk parameter is estimated (close) to zero, it is still possible to identify model improvements by fixing the diffusion to a value that allowed regular and sufficient updating. In the presented example this is fixed such that the diffusion is comparable to the diffusion of other states of the system.

All suggested model improvements was tested by means of information criteria or likelihood ratio testing. The first part of the identification, based on random walk state estimation for identification and information criteria for selection, resulted in large reductions of the diffusion and the observation variance, implying that the deterministic skeleton dictated by the drift term, provides a better description of the model as the complexity of the model increase. The validation step is based on simulation studies, which also pinpointed model deficiencies and suggested further model extensions for both drift and diffusion. These model extensions lead to significant improvements of the likelihood, but more importantly to a stable simulation model and reductions of the simulation variance while improving the simulation performance (in terms of quantile skill score).

The diffusion term is shown to be important for both short-term (likelihood estimation) and long-term (simulations) dynamics of the system. In the present context we are limited by the ability to find solutions for the Lamperti transform and an explicit inverse transformation. This excludes a large class of diffusion processes, where mass balance or partial mass balance is taken into account in the formulation of the diffusion matrix. The model development did however demonstrate that the correlation coefficient is not an important parameter in the model.

The simplicity of the model example implies that most parameters are well determined in a statistical sense. For more complicated models (or models with less information provided by the observations) it might be appropriate to include prior knowledge in the optimisation. The statistical software used here, does include the possibility of including prior knowledge in the estimation through the Maximum A Posterior (MAP) procedure described in Kristensen et al. (2004a). Such prior knowledge could be obtained from literature studies or from site specific experiments.

References

Aït-Sahalia, Y. (2008) Closed-form likelihood expansions for multivariate diffusions. *The Annals of Statistics*, **32**, (2), 906-937

Baadsgaard, M., Nielsen, J. N., Spliid, H., Madsen, H., and Preisel, M. (1997) Estimation in stochastic differential equations with state dependent diffusion term. *SYSID '97 - 11th IFAC symposium on system identification, IFAC*

Bak, J., Nielsen, H. Aa., and Madsen, H. (1999) Goodness of fit od stochastic differential equations. *Symposium i Anvendt Statistik, Copenhagen*, 341-346

Bartell, S. M., Lefebvre, G., Kaminski, G., Carreau, M., and Campbell, K. R. (1999) An ecosystem model for assessing ecological risks in Québec river, lakes and reservoirs. *Ecological Modelling*, **124**, 43-67

Fasham, M. J. R, Ducklow, H. W., and McKelvie S. M. (1990) A nitrogen-based model of plankton dynamics in the oceanic mixed layer. *Journal of Marine Research*, **58**, 591-639

Fisher, T. R., Peele, E. R., Ammerman, J. W., and Harding, L.W. (1992) Nutrient limitation of phytoplankton in Chesapeake Bay. *Marine Ecology Progress Series* **82**, 51-63.

Forman, J., and Sørensen, M. (2008) The Pearson Diffusion: A Class of Statistically Tractable Diffusion Processes. *Scandinavian Journal of Statics*, **35**, 438-465

Gard, T. C. (1988) Introduction to stochastic differential equations. *Marcel Dekker, Inc.*, New York.

Gneiting, T., and Raftery, A. E. (2007) Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, **102**(477), 359-378

Hastie, T. J., and Tibshirani, R. J. (1990) Generalized additive models. *Monographs on Statistics and Applied Probability*, **43**, Chapman & Hall/CRC

Hastie, T., Tibshirani, R., and Friedman, J. (2001) The Elements of Statistical Learning - Data Mining, Inference and Prediction. *Springer Series in Statistics, Springer New York.*

Iacus, S. M. (2008) Simulation and Inference for Stochastic Differential Equations - With R Examples. *Springer Series in Statistics, Springer New York.*

Jazwinski, A. H. (1970) Stochastic Processes and Filtering Theory. *Academic Press, New York, ASU.*

Karatzas, I., and Shreve S. E. (1991) Brownian Motion and Stochastic Calculus - Second edition. *Springer New York.*

Klebaner, F. C. (2005) Introduction to Stochastic Calculus with Applications. *Imperial College Press, London.*

Kloden, P., and Platen, E. (1999) Numerical solutions of Stochastic Differential Equations. *Springer-Verlag, Berlin.*

Koenker, R., and Bassett, G. J. (1978) Regression Quantile. *Econometrica*, **46**(1), (33-50).

Kristensen, N. R., and Madsen, H. (2003) Continuous time stochastic modelling - CTSM 2.3 - Mathematics Guide. *Technical University of Denmark*

Kristensen, N. R., Madsen, H., and Jørgensen, S. B. (2004a) Parameter estimation in stochastic grey-box models. *Automatica*, **40**, 225-237

Kristensen, N. R., Madsen, H., and Jørgensen, S. B. (2004b) A method for systematic improvement of stochastic grey-box models. *Computers & Chemical Engineering*, **28**, 1431-1449

Luschgy, H., and Pagés, G. (2006) Functional quantization of a class of Brownian diffusions: A constructive approach. *Stochastic Processes and their Applications* **116***, 310-336*

Madsen, H. (2008) *Time Series Analysis.* Chapman & Hall/CRC

Madsen, H., Holst, J., and Thyregod P. (1987) A Continuous Time Model for the Variations of Air Temperature. *10. Conference on Probability and Statistics in Atmospheric Science, American Meteorological Society,* pp. 52-58. Edmonton

Møller, J. K., Nielsen, H. Aa., and Madsen, H. (2008). Time-adaptive quantile regression. *Computational Statistics & Data Analasis*, **52**, 1292-1303

Nielsen, H. Aa., and Madsen, H. (2001) A Generalization of some Classical Time Series Tools. *Computational Statistics & Data Analysis* **37**, 13-31

Pedersen, M. W., Righton, D., Thygesen, U.H., Andersen K., and Madsen, H. (2008). Geolocation of North Sea cod using Hidden Markov Models and behavioral switching. *Canadian Journal of Fisheries and Aquatic Sciences*, Vol. 65, pp. 2367-2377

Pinson, P., Nielsen, Aa. H., Møller, J. K., and Madsen, H. (2007) Non-parametric Probabilistic Forecast of Wind Power: Required Properties and Evaluation. *Wind Energy*, **10**, pp. 497-516

Tornøe, C. W., Jacobsen, J., Pedersen, O., Hansen, T., and Madsen, H. (2004) Grey-box Modelling of Pharmacokinetic/Pharmacodynamic Systems. *J. Pharmacokinet. Pharmacodyn*, Vol. 31, pp. 401-417

Øksendal, B. (2003) Stochastic Differential Equati¡ons - An Introduction with Applications, Sixth edition. *Springer-verlag, Berlin*

Paper G

Development of a Restricted State Space Stochastic Differential Equation Model for Bacterial Growth in Rich Media

---

Authors:
J.K. Møller, K.R. Philipsen, L.E. Christiansen, and H. Madsen.

Submitted to:
Journal of Theoritical Biology, (2010).

Development of a Restricted State Space Stochastic Differential Equation Model for Bacterial Growth in Rich Media

Jan Kloppenborg Møller[1], Kirsten Riber Philipsen[1], Lasse Engbo Christiansen[1], and Henrik Madsen[1]

Abstact

In the present study bacterial growth in a rich media is analysed in a Stochastic Differential Equation (SDE) framework. It is demonstrated that the SDE formulation and smoothen state estimates provide a systematic framework for data driven model improvements, using random walk hidden states. Bacterial growth is limited by the available substrate and the inclusion of diffusion must obey this natural restriction. By inclusion of a modified logistic diffusion term it is possible to introduce a diffusion term flexible enough to capture both the growth phase and the stationary phase, while concentration is restricted to the natural state space (substate and *bacteria non-negative). The case considered is growth for Salmonella and Enterococcus in a rich media. It is found that a hidden state is necessary to capture the lag phase of growth, and that a flexible logistic diffusion term is needed to capture the random behaviour of the growth model. Further, we find that the Monod effect is not needed to capture the dynamics of bacterial growth in the presented data.*

Key Words: Stochastic Differential Equations, Logistic Diffusion, Extended Kalman Filter, Bacterial growth, Monod growth, Rich media, Enterococcus, Salmonella.

1 Introduction

Stochastic Differential Equations (SDEs) has shown great potential for modelling in various areas of applications, e.g. Tornøe et al. (2004), Pedersen et al. (2008) and Madsen et al. (1987). In general such models are based on formulations where a constant diffusion term is added to a non-linear drift term,

[1]DTU Informatics, Richard Pedersens Plads, Technical University of Denmark Building 321, DK-2800 Lyngby, Denmark.

which in turn is derived from known theory. In many cases this is a fruitful approach, but for systems with a restricted state space this might not be the best approach, since additive diffusion will allow the system to leave the physically feasible state space, called the natural state space. A main theme of this paper is to provide a proper description of diffusion in restricted bacterial growth models.

It is well known that SDEs give more correct state estimation because filtering techniques allow the information given by observations to be incorporated into the state estimate. In addition to improved state estimation a SDE approach provide a framework for identification of model deficiencies (Kristensen et al., 2004b). In this paper the method is applied to bacterial growth data and allow us to argue for possible model improvements.

Parameter estimation in SDEs is a difficult task and approximate methods have to be applied for complex systems. In this study we apply a statistical method based on the Extended Kalman Filter (EKF) (Jazwinski, 1970), implemented in the statistical software CTSM[1] (Kristensen & Madsen, 2003; Kristensen et al., 2004a). CTSM has proven powerful in many application (e.g. Philipsen et al. (2010a)); it is however not possible to include state dependent diffusion in CTSM. One way to overcome this restriction is by including some input in the description of the diffusion term (e.g. state observations (Philipsen et al., 2010a)). A more correct description can be obtained by applying Lamperti type transformations (e.g. Iacus (2008)) to obtain a system with state independent diffusion from a system with state dependent diffusion. Under the restriction that Lamperti transformations are available, different parametric representations of the diffusion term are derived and evaluated with respect to a case study of bacterial growth. In particular a novel modified logistic diffusion term will be introduced and evaluated.

The case study in this paper is a bacterial growth model, and an adequate description of bacterial growth kinetics, i.e. the relationship between bacterial growth and substrate concentration, is important for many applications in microbiology, for instance for fermentation processes (Kompala et al., 1984; Patnaik, 1999). The Monod equation was the first suggestion for a mathematical description of the growth curve. It has been extensively discussed since its introduction in 1949 (Monod, 1949). When originally proposed, it seemed to be a "convenient and logical" (Monod, 1949) choice for a mathematical expression to fit the growth curve. Several attempts have been made to formulate a mechanistic background for the Monod equation (Liu, 2006; Lobry et al., 1992). Even though the Monod equation is a good description of the growth on a single substrate, it fails to model growth in rich media (Bajpai-Dikshit et al.,

[1] www2.imm.dtu.dk/~ctsm

2003). Therefore several attempts have been made to find another equation for this growth (Doshi et al., 1997). According to Kovárová-Kovar and Egli (1998) these studies can be divided into three methods i) Extending the Monod model with additional constants, ii) Developing an empirical or mechanistic model from kinetic concepts, iii) Describing how the Monod growth parameters are influenced by physiochemical factors.

A problem with determining other expressions for the growth rate has been the lack of high quality reproducible data which relates the growth rate to the substrate concentration (Kovárová-Kovar and Egli, 1998). The method proposed here makes it possible to extract these data from bioscreen measurements, and thus limit the time and resources used on experiments significantly.

The transition from modelling using Ordinary Differential Equations (ODEs) to Stochastic Differential Equations (SDEs) paves the way for many strong statistical tools for model development and inference (Kristensen et al., 2004a). In this paper a model development method based on SDEs (and the applicability of the Lamperti transform) will be introduced to examine and subsequently describe the relation between growth rate and substrate concentration. The use of SDEs enable the separation of measurement noise and system noise, which will be used in the method. First the SDE framework will be introduced, followed by a short presentation of data. Then a thorough analysis of the diffusion term is perform to determine the best way to introduce diffusion in the model. In the last part we developed the drift term and present a simulation study of the final model.

2 Methodology

The systematic framework in this study is a continuous-discrete time stochastic state-space formulation

$$d\boldsymbol{X}_t = \boldsymbol{f}(\boldsymbol{X}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})dt + \boldsymbol{\sigma}(\boldsymbol{X}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})d\boldsymbol{w}_t \tag{1}$$

$$\boldsymbol{Y}_k = \boldsymbol{h}(\boldsymbol{X}_k, \boldsymbol{u}_k, t_k, \boldsymbol{\theta}) + \boldsymbol{e}_k \ , \tag{2}$$

where Eq. (1) is the continuous time system equation and Eq. (2) is the discrete time observation equation. $\boldsymbol{X}_t$ is a vector of state variables, $\boldsymbol{Y}_k$ is a vector of measured output variables at times t_k, $\boldsymbol{u}_t$ is a vector of known input variables, $\boldsymbol{e}_k$ is an l-dimensional white noise process with $\boldsymbol{e}_k \in \mathcal{N}(0, \boldsymbol{S}(\boldsymbol{u}_k, t_k, \boldsymbol{\theta}))$, $\boldsymbol{w}_t$ is a standard Wiener process with zero mean and independent Gaussian time increments, and $\boldsymbol{\sigma}(\boldsymbol{X}_t, \boldsymbol{u}_t, t, \boldsymbol{\theta})$ is the diffusion coefficient . The first part of the system equation is called the drift term and the second part is called the

diffusion term. Everywhere in this article the Itô interpretation of the SDEs is used (Øksendal, 2003).

2.1 Estimation

The estimation procedure is a maximum likelihood procedure in which the EKF is used to evaluate the likelihood function. A full account is available in Kristensen & Madsen (2003), and we will not present the details here. We will however give a few remarks on the output from the software and the restrictions of the procedure.

In addition to the parameter estimates and the parameter covariance matrix (estimated by the inverse Hessian, e.g. Madsen (2008)), the implementation allows us to calculate k-step predictions of both the state and the observations, and the smoothened state (state estimate given all observations).

Of special interest is the one-step prediction, the standardised residuals for one dimensional observations are given by

$$r_k = \frac{Y_k - \hat{Y}_{k|k-1}}{\sqrt{\Sigma^{yy}_{k|k-1}}}, \tag{3}$$

where $\hat{Y}_{k|k-1}$ is the one-step prediction of the observations and $\Sigma^{yy}_{k|k-1}$ is the variance of the one-step prediction. If the model has captured the mechanistic behaviour of data then r_k is a white noise process. If further the sampling intervals are equidistant then r_k can be analysed by traditional residual analysis, like the autocorrelation function and the partial autocorrelation function.

In addition to the improved state estimates, smoothen state estimates of a modified system can help to identify model deficiencies (Kristensen et al., 2004b). A suspicion that the model is not sufficient to describe the variation in data, can be analysed by considering smoothen state estimates of the extended state space

$$d\begin{bmatrix} \boldsymbol{X}_t \\ \theta^i_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{f}(\boldsymbol{X}_t, \boldsymbol{u}_t, t, \theta^i_t, \widetilde{\boldsymbol{\theta}}) \\ 0 \end{bmatrix} dt + \begin{bmatrix} \boldsymbol{\sigma}(\boldsymbol{X}_t, \boldsymbol{u}_t, t, \widetilde{\boldsymbol{\theta}}) & \mathbf{0} \\ \mathbf{0} & \sigma_\theta \end{bmatrix} d\boldsymbol{w}_t \,, \tag{4}$$

where θ^i_t is the parameter to be investigated and $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta} \setminus \theta^i_t$. By plotting the smoothen state of θ^i_t as a function of time and possible covariates model improvements can be identified.

2.2 Transformation of the state space

As mentioned above it is not possible to include processes with state dependent diffusion in CTSM (i.e. the SDE models are restricted to the form $\boldsymbol{\sigma}(\cdot) = \boldsymbol{\sigma}(\boldsymbol{u}_t, t, \boldsymbol{\theta})$). For one dimensional diffusion processes with a state dependent diffusion term it is, however, always possible to find a transformation of the state space such that the diffusion is independent of the states (e.g. Baadsgaard et al. (1997)). The transformation is often referred to as the Lamperti transform (Iacus, 2008)

$$Z_t = \psi(X_t, \cdot) = \int \frac{d\xi}{\sigma(\xi, \cdot)}\bigg|_{\xi = X_t}, \tag{5}$$

and by Itô's lemma (e.g. Øksendal (2003)), Z_t will again be an Itô process, given by (Luschgy and Pagés (2006))

$$dZ_t = \left(\psi_t(\psi^{-1}(Z_t, \cdot), t) + \frac{f(\psi^{-1}(Z_t, \cdot), \cdot)}{\sigma(\psi^{-1}(Z_t, \cdot), \cdot)} dt - \frac{1}{2}\sigma_x(\psi^{-1}(Z_t, \cdot), \cdot) \right) dt + dw, \tag{6}$$

where subscript refer to partial differentiation. The SDE description (6) have state independent diffusion. The practical applicability of the Lamperti transform is limited to cases where we can find an explicit inverse ($\psi^{-1}(Z_t, \cdot)$). The existence of a Lamperti transform with an explicit inverse is the basis of the presented analysis.

3 Data

Optical density measurements for the growth of a *Salmonella* strain and a *Enterococcus* strain growing in BHI media are available for the analysis. For each bacteria culture a 9-fold dilution as well as the non-diluted strain are measured in duplicates. The measurements are made for 40 hours with a sampling interval of 20 minutes in a bioscreen (Microbiology Reader Bioscreen C) at 16°C under continuous shaking.

The OD measurements are preliminary corrected for background broth measurements and subsequently a correction is made for the shadow effect for high OD values by the relation (Philipsen et al., 2010b)

$$\mathrm{OD}_{\mathrm{corr}} = -\frac{1}{b} \log\left(1 - \frac{\mathrm{OD}_{\mathrm{meas}}}{a}\right). \tag{7}$$

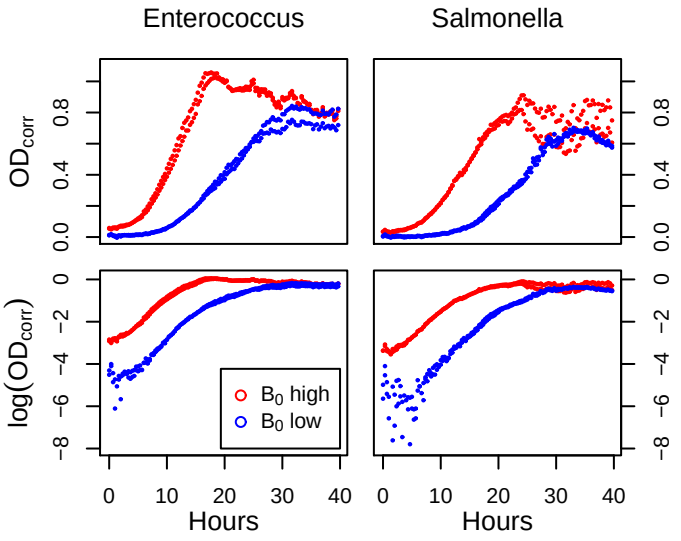

Figure 1: Untransformed optical density measurements (top row) and log-transformed optical density measurements (bottom row).

The constants a and b are found by fitting the relation to results from an experiment with *Salmonella* for which the OD for different concentrations has been measured. It is assumed that the same parameters can be used for *Enterococcus*. The OD measurements after corrections, $\mathrm{OD}_{\mathrm{corr}}$, are shown in Figure 1.

The models that are developed in this work are not suited for modelling the transition to the stationary phase (or the stationary phase itself). The transition is however needed in order to be able to estimate the stoichiometric constant η (see Section 4 below). It is therefore decided to include all data-points from experiments and the aim is to capture the stationary phase by the diffusion term.

Data show little variation in the beginning of the data set and strong variation at the end of the dataset for the untransformed data (Figure 1, top panels). This behaviour indicates that data should be transformed before parameter estimation. log-transformed data does however show strong variation in the beginning and only small variations at the end of the dataset (Figure 1, bottom panels), indicating that we should choose something in between, and here we choose the Box-Cox transformation (Madsen (2008))

$$Y_t^{(p)} = \begin{cases} (Y_t^p - 1)/p & p \neq 0 \\ \log(Y_t) & p = 0 \end{cases}, \tag{8}$$

the choice of p is based on the initial estimation model and is discussed in Section 5.

4 A minimal stochastic growth model

A bacterial growth model should include mechanisms to ensure that the growth rate is zero when the substrate is depleted and that substrate and bacteria levels are both positive at all times. While these restrictions are simple in an ODE setting, it is more complicated in a SDE setting, hence the diffusion term must be handled with care. A first general formulation is

$$d\begin{bmatrix} B_t \\ S_t \end{bmatrix} = \begin{bmatrix} \widetilde{\mu}(S_t, B_t, \boldsymbol{\theta})B_t \\ -\eta\widetilde{\mu}(S_t, B_t, \boldsymbol{\theta})B_t \end{bmatrix} dt + \begin{bmatrix} \tilde{\sigma}_B(S_t, B_t, \boldsymbol{\theta}) \\ -\eta\tilde{\sigma}_B(S_t, B_t, \boldsymbol{\theta}) \end{bmatrix} dw, \tag{9}$$

where η is a stoichiometric conversion from bacteria units to substrate units, and $\widetilde{\mu}(\cdot)$ and $\widetilde{\sigma}_B(\cdot)$ are functions to be determined.

Now consider $T_t = S_t + \eta B_t$. Clearly T_t is constant ($dT_t = 0$) and S_t can be expressed as

$$S_t = T_t - \eta B_t = S_0 + \eta(B_0 - B_t). \tag{10}$$

and (9) reduces to the one-dimensional diffusion process

$$dB_t = \widetilde{\mu}(S_t, B_t, \theta)B_t dt + \widetilde{\sigma}_B(S_t, B_t, \theta)dw, \tag{11}$$

with S_t is given by (10).

S_t is implemented as standardised substrate such that $T_t = T_0 = 1$. The simplest model for the drift term that ensure substrate *and* bacteria concentration to be above zero at all times is $\widetilde{\mu}(\cdot) = \mu S_t B_t$, (with μ constant). It is tempting to include diffusion as proportional to the state, i.e. $\tilde{\sigma_B}(S_t, B_t, \theta) = \sigma_B B_t$, with σ_B constant (corresponding to variance proportional to B_t^2). Such a formulation does, however, not include a mechanism to kill the diffusion as S_t approaches zero and will therefore hold a positive probability of $\eta B_t > T_0$ (implying $S_t < 0$), which is clearly a violation of the natural constraints for the system.

To include the mass balance in the diffusion term, we consider a logistic diffusion term (Schurz , 2007), i.e. $\tilde{\sigma_B}$ is modelled as

$$\widetilde{\sigma}_B(S_t, B_t, \theta) = \sigma_B B_t^{\alpha} S_t^{\beta}, \tag{12}$$

where σ_B, α and β are real constants. Schurz (2007) analyses a class of logistic SDEs (including a logistic drift term) and prove that for $\alpha \geq 1$ and $\beta \geq 1$ the

logistic SDE will stay within the boundary defined by the logistic ODE. Others (e.g. Iacus (2008)) suggests $\alpha = \beta = \frac{1}{2}$, this formulation is however (as shown below) not suited for estimation with the EKF. The Lamperti-transform (for $\alpha = \beta = \frac{1}{2}$) is given by

$$Z_t^B = \int \frac{d\xi}{\sqrt{\xi(T_0 - \eta\xi)}}\bigg|_{\xi=B_t} = -\frac{1}{\sqrt{\eta}} \arctan\left(\frac{T_0 - 2\eta B_t}{2\sqrt{\eta B_t(T_0 - \eta B_t)}}\right), \tag{13}$$

and the resulting Itô process is

$$dZ_t^B = \left(\mu\sqrt{S_t B_t} - \frac{\sigma_B^2(T_0 - 2\eta B_t)}{2\sqrt{B_t S_t}}\right) dt + \sigma_B dw_t, \tag{14}$$

the state space of Z_t^B is the interval $\left[-\frac{\pi}{2\sqrt{\eta}}, \frac{\pi}{2\sqrt{\eta}}\right]$, and the boundaries are reached at $B_t = 0$ and $S_t = 0$, respectively. Considering the asymptotic behaviour we get

$$\lim_{B_t \to 0} dZ_t^B = \lim_{B_t \to 0} \left(\left(\mu\sqrt{S_t B_t} - \frac{\sigma_B^2(T_0 - 2\eta B_t)}{2\sqrt{B_t S_t}}\right) dt + \sigma_B dw_t\right) \tag{15}$$

$$= \lim_{B_t \to 0} \left(\left(-\frac{\sigma_B^2 T_0}{2\sqrt{B_t S_t}}\right) dt + \sigma_B dw_t\right) \tag{16}$$

$$= -\infty dt \quad a.s., \tag{17}$$

by similar calculations we get,

$$\lim_{S_t \to 0} dZ_t^B = \infty dt, \quad a.s. \tag{18}$$

The problem is that the process holds a positive probability of bacteria extinction and a positive probability of all the substrate being depleted. Even though it is possible for bacteria to deplete the limiting source complete, comparing such a hypothesis with data seems unreasonable (Figure 1, top row), since there is a clear transition to a stationary phase with continuous variations of bacteria concentration. There is also a computational issue here, if Z_t^B leaves the state space we cannot evaluate the state estimate and the likelihood. It might seem that this is a weakness stemming from the combination of the Lamperti transform and the EKF. At the extreme $S_t \to 0$ we can view (14) as a special case of the (transformed) Feller diffusion (Iacus, 2008). The parameters of this particular Feller diffusion (additive constant drift equal zero) will lead to a transition *pdf*, which tends to infinity at $S_t \to 0$ (Feller, 1951). Direct evaluation of the Fokker-Planck equation (Gard, 1988), will therefore also be difficult or impossible, and even simulation based methods will have difficulties evaluating the transition densities.

In the following Section 4.1 a minimal model with the desired characteristics is defined and in Section 6 we define a more flexible diffusion term, which gives a better representation of data.

4.1 Initial estimation model

Based on the reasoning above the minimal drift term ($\widetilde{\mu}(\cdot) = \mu B_t S_t$) and a logistic diffusion with $\alpha = \beta = 1$ is chosen for the initial model which is then given by

$$dB_t = \mu S_t B_t dt + \sigma_B S_t B_t dw_t, \tag{19}$$

where $S_t = T_0 - \eta B_t$.

In the following we will develop the model assuming that μ is not constant, but we will use (19) to determined an appropriate transformation of data and develop the diffusion term further before we turn to the drift term. Formulated in terms of B_t (19) is given by

$$dB_t = \mu(T_0 - \eta B_t) B_t dt + \sigma_B (T_0 - \eta B_t) B_t dw_t. \tag{20}$$

The Lamperti transform is

$$Z_t^B = \int \frac{d\xi}{\xi(T_0 - \eta\xi)}\bigg|_{\xi = B_t} = \frac{1}{T_0} \log\left(\frac{B_t}{T_0 - \eta B_t}\right), \tag{21}$$

with the inverse

$$B_t = T_0 \frac{e^{Z_t^B T_0}}{1 + \eta e^{Z_t^B T_0}}, \tag{22}$$

and the transformed system is given by

$$dZ_t^B = \left(\mu + \frac{1}{2}\sigma_B^2 (2\eta B_t - T_0)\right) dt + \sigma_B dw_t. \tag{23}$$

The system (23) is intentionally not transformed to unit diffusion, because we want to be able to control the initial conditions independently of the parameters (see Section 5).

The state space of Z_t^B is the entire real axis, and for $B_t \to 0$ and $\eta B_t \to T_0$ the drift term in dZ_t^B is a constant and for finite t, Z_t^B is finite *a.s.*. We will therefore use the formulation (23) as a starting point for the model development.

5 Estimation issues

As mentioned earlier there are two experiments for each bacteria, one with high starting concentration and one with low starting concentration, and two

replications of each experiment. Low starting concentration is reported to be a 9 fold dilution of the high starting concentration, this relations is however not evident from the data, and it was decided to estimate the dilution.

The estimation of dilution is performed by starting the process at time $t = -1min$, with B_{-1} close to zero (in practice $Z^B_{-1} = -10$) and letting the estimation procedure estimate the starting concentration by integrating a modified growth model

$$dB_t = (\mu S_t B_t + C_1 u^h_t + C_2 u^l_t)dt + \sigma_B S_t B_t dw_t, \tag{24}$$

where $u^l_t = 1$ for $t < 0$ and low starting concentration and zero otherwise, and $u^h_t = 1$ for $t < 0$ and high starting concentration and zero otherwise. In the following we will suppress these inputs, but note that they will be present throughout the analysis.

To be able to choose the best transformation of data, one-step predictions from the model formulation (24) was calculated for different choices of p in the Box-Cox transformation (8), clearly an assumption of constant variance is not reasonable for p equal 0.25 and 0.5 (Figure 2), while for $p = 0.75$ the assumption seems reasonable. It was therefore decided to use a Box-Cox transformation with $p = 0.75$ for the data analysis, and thus the observation equation used in the remainder of this paper is

$$\frac{4}{3}(OD^{3/4}_{\text{corr},k} - 1) = \frac{4}{3}(B^{3/4}_{t_k} - 1) + e_k, \tag{25}$$

with $e_k \sim N(0, s^2_y)$.

6 Model development

The shape of the diffusion term determines the random behaviour of the bacterial growth process. The drift terms of the models that will be analysed here are well suited for describing the growth part of the process, while the transition to the stationary phase is not included. The stationary phase should therefore be captured by the diffusion term. However, the diffusion term introduced above has maximum at $\eta B_t = \frac{1}{2}$, while the transition to the stationary phase should be close to 1, which is where the growth stops. It is therefore reasonable to assume that the maximum of the diffusion is assumed at a value in the interval $(\frac{1}{2}, 1)$.

In order to introduce this kind of random behaviour we propose the a modified

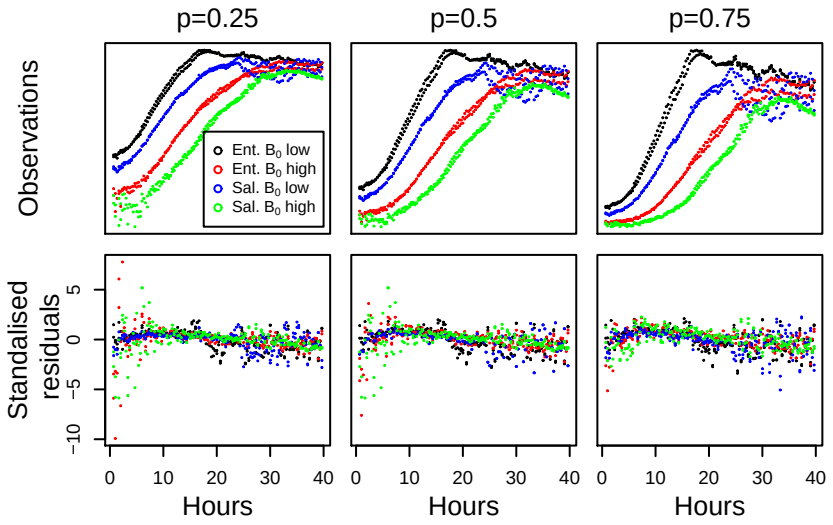

Figure 2: Box-Cox transformed data (top row) and standardised residuals from the model formulation Eq. (24) for different values of p.

logistic diffusion term for modelling of random bacterial diffusion

$$\widetilde{\sigma}_B(B_t, S_t, \theta) = \sigma_B B_t (T_0 - (\eta B_t)^\gamma), \tag{26}$$

where γ is a positive constant, the Lamperti transform for this diffusion is

$$Z_t^B = \int \frac{d\xi}{\xi(T_0 - (\eta\xi)^\gamma)}\bigg|_{\xi=B_t} = \frac{1}{\gamma T_0} \log\left(\frac{(\eta B)^\gamma}{T_0 - (\eta B)^\gamma}\right), \tag{27}$$

with the inverse

$$B_t = \frac{T_0}{\eta} \frac{e^{T_0 Z_t^B}}{(1 + e^{\gamma T_0 Z_t^B})^{1/\gamma}}. \tag{28}$$

The state space of Z_t^B is the real line and the Itô process for Z_t^B is given by

$$dZ_t^B = \left(\mu \frac{T_0 - \eta B_t}{T_0 - (\eta B_t)^\gamma} - \frac{1}{2}\sigma_B^2 (T_0 - (\gamma + 1)(\eta B_t)^\gamma)\right) dt + \sigma_B dw_t, \tag{29}$$

where B_t is given by the inverse transformation (28).

Applying L'Hôpital's rule (with $T_0 = 1$) we get

$$\lim_{\eta B \to T_0} dZ_t^B = \left(\frac{\mu}{\gamma T_0^\gamma} - \frac{1}{2}\sigma_B^2 (T_0 - (\gamma + 1) T_0^\gamma)\right) dt + \sigma_B dw_t, \tag{30}$$

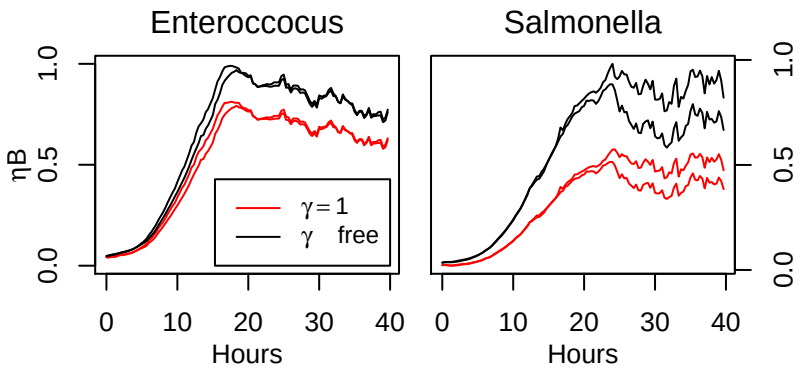

Figure 3: Smoothen states of the bacteria concentration measured in substrate units, for Model 0 for both $\gamma = 1$ and γ estimated as a parameter. Due to low observation noise observations are almost consistent with the smoothen state and therefore not shown.

implying that Z_t^B is finite $a.s.$ for finite t.

The introduction of γ improved the likelihood significantly (Table 1) for both Enterococcus and Salmonella. The stoichiometric constant, η, also changes significantly (Table 1). The consequence of the small values of η in the initial model set up (19), compared to the values of η estimated by the modified logistic diffusion (26), is that transition to the stationary phase is estimated to be close to the maximum of the diffusion term (i.e. $\frac{1}{2}$ when $\gamma = 1$, Figure 3).

Table 1: Estimation results for models with $\gamma = 1$ and γ part of the estimation.

	γ (s.d.)	η (s.d.)	DF	log(L)	AIC	BIC
Enterocccous	**1.0**(NA)	**0.77**(0.024)	6	1213.7	-2416	-2391
Enterocccous	**21.0**(6.5)	**0.94**(0.004)	7	1261.7	-2510	-2481
Salmonella	**1.0**(NA)	**0.63**(0.090)	6	1111.7	-2212	-2187
Salmonella	**68.9**(9.9)	**1.08**(0.002)	7	1163.9	-2314	-2285

Based on the likelihood results presented in Table 1 and the conventional knowledge that bacteria will consume the available substrate, the model analysis in the following include γ, and we refer to the model (24) with the diffusion (26) as Model 0. The large values of γ imply that the diffusion is close to state proportional diffusion, except very close to upper bound for bacteria concentration, where the diffusion is killed of very rapidly. The key difference is however that the presented model guarantees that $S_t > 0$, whereas state proportional

(or constant) diffusion will allow $S_t < 0$.

6.1 Non-parametric identification of the growth process

By formulating μ as a pure random walk hidden state adapted to data it is possible to analyse the growth rate and suggest further model improvements. The model formulation in the transformed domain is

$$d\begin{bmatrix} Z_t^B \\ \mu_t \end{bmatrix} = \begin{bmatrix} \mu_t \frac{T_0-\eta B_t}{T_0-(\eta B_t)^\gamma} - \frac{1}{2}\sigma_B^2(T_0-(\gamma+1)\eta B_t) \\ 0 \end{bmatrix} dt + \begin{bmatrix} \sigma_B & 0 \\ 0 & \sigma_\mu \end{bmatrix} d\boldsymbol{w}_t. \tag{31}$$

where B_t is given by the inverse transformation (28).

The smoothen state (state estimate given all observations) of μ_t can now be analysed to get an idea of the possible model improvements. The shape of the μ_t-curves varies substantially between starting concentrations and bacteria (Figure 4). It is also evident that simple functions of time or substrate (or bacteria) will not be sufficient to capture these shapes. The growth parameter is linearly dependent on the substrate in the growth phase, implying a second order interaction in the growth-rate of bacteria. Also the autocorrelation function of the residuals from Model 0 (Figure 5) has many significant values. Such behaviour suggest that an additional state could improve the model.

Following Bajpai-Dikshit et al. (2003) a possible extension of Model 0 is

$$\mu_t = vE_t, \tag{32}$$

with

$$\frac{dE_t}{dt} = \frac{v+\beta}{\kappa+S_t}S_t - E_t\frac{d\log(B_t)}{dt} - \beta E_t, \tag{33}$$

where v is the maximum bacterial growth rate, β is the first-order degradation of proteins inside the cell, see Bajpai-Dikshit et al. (2003) for more specific details of the formulation.

To keep derivations simple and preserve flexibility of the enzyme diffusion, we consider only the deterministic part of dB_t in the derivations of dE_t, for special cases of the enzyme diffusion a full stochastic inclusion of $d\log(B_t)$ is however possible (see A). The deterministic formulation of dE_t is

$$dE_t = \left((v+\beta)\lambda_t - E_t^2\lambda_t - \beta E_t\right)dt, \tag{34}$$

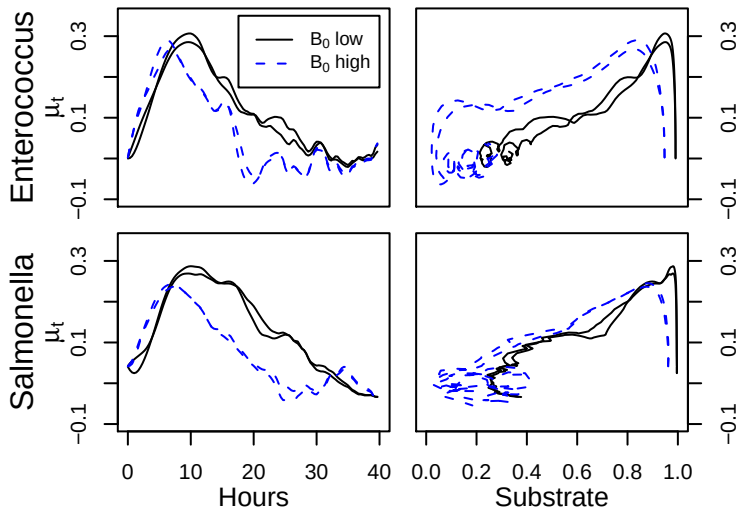

Figure 4: Growth rates estimated from the model presented in Eq. (31).

with $\lambda_t = \frac{S_t}{\kappa + S_t}$. Following the modelling procedure in Bajpai-Dikshit et al. (2003) the bacteria growth rate should also follow the Monod term given in λ_t. Adding a diffusion term to dE_t give the full SDE description

$$d\begin{bmatrix} B_t \\ E_t \end{bmatrix} = \begin{bmatrix} E_t v \lambda_t B_t \\ (v+\beta)\lambda_t - E_t^2 v \lambda_t - \beta E_t \end{bmatrix} dt + \begin{bmatrix} \sigma_B B_t (T_0 - (\eta B_t)^\gamma) & 0 \\ 0 & \widetilde{\sigma}_E(E_t) \end{bmatrix} d\boldsymbol{w}. \tag{35}$$

Equation (34) describes the relative (to the maximum) level of enzymes and E_t should therefore be restricted to the interval $[0, 1]$. Even though this is theoretically the case, there is no break down of the dynamics of the equation even if E_t exceeds 1 (but v cannot be interpreted as maximum growth-rate). The dynamics does however break down if E_t is allowed to be below 0. The remainder of the paper presents results for different choices of λ_t and $\widetilde{\sigma}_E(E_t)$.

7 Estimation results

The model presented in Eq. (35) is estimated for different specifications of λ_t and $\widetilde{\sigma}_E(E_t)$ (Table 2). The combination are λ_t equal S_t or including Monod

growth, and enzyme diffusion proportional to the enzyme level ($\widetilde{\sigma}_E(E_t) = \sigma_E E_t$) or enzyme diffusion which ensures enzyme levels to stay within the interval $(0, 1)$ ($\widetilde{\sigma}_E(E_t) = \sigma_E E_t(1 - E_t)$). The likelihood results are summarised in Table 2, while parameter estimates are given in Table 3 and 4.

Table 2: Likelihood table for Enterococcus and Salmonella, Δ_0^{AIC} and Δ_0^{BIC} refer to the difference in AIC and BIC compared to Model 0.

	λ_t	$\sigma_E(\cdot)$	log(L)	DF	Δ_0^{AIC}	Δ_0^{BIC}
		Enterococcus				
Model 1	S_t	$\sigma_E E_t$	1340.6	9	-153.7	-145.4
Model 2	$\frac{S_t}{\kappa+S_t}$	$\sigma_E E_t$	1340.6	10	-151.7	-139.2
Model 3	S_t	$\sigma_E E_t(1 - E_t)$	1334.3	9	-141.2	-132.9
Model 4	$\frac{S_t}{\kappa+S_t}$	$\sigma_E E_t(1 - E_t)$	1348.7	10	-168.0	-155.5
Model 5	S_t	$\sigma_E E_t(1 - E_t)$	1348.6	9	-169.9	-161.5
		Salmonella				
Model 1	S_t	$\sigma_E E_t$	1194.2	9	-56.8	-48.4
Model 2	$\frac{S_t}{\kappa+S_t}$	$\sigma_E E_t$	1194.2	10	-54.7	-42.3
Model 3	S_t	$\sigma_E E_t(1 - E_t)$	1201.8	9	-71.8	-63.5
Model 4	$\frac{S_t}{\kappa+S_t}$	$\sigma_E E_t(1 - E_t)$	1202.8	10	-71.9	-59.4
Model 5	S_t	$\sigma_E E_t(1 - E_t)$	1204.0	9	-76.3	-68.0

The models are not nested and likelihood ratio testing cannot be applied, the evaluation is therefore based on AIC and BIC values. The inclusion of an enzyme state without the Monod term and with state proportional enzyme diffusion (Model 1) gives highly significant improvements in AIC and BIC for both Enterococcus and Salmonella and all parameters are significant (except enzyme diffusion for Salmonella) for both datasets.

The further inclusion of the Monod term (Model 2) does not give further improvement of AIC and BIC for any of the two datasets (Table 2). The parameters v and κ are both estimated to very high (and insignificant) values (Tables 3 and 4), and there is evidence that the Monod term should not be included.

Removing the Monod term and including a logistic diffusion term for the Enzyme level (Model 3) give improvements in both AIC and BIC for Salmonella, while AIC and BIC increase for Enterococcus (Table 2). The parameter estimate of σ_E also changes significantly (Tables 3 and 4), reflecting that E_t is well determined by the deterministic equation when E_t is close to 0 or 1, while it is dominated by diffusion for E_t away from the boundary.

The further inclusion of the Monod term (Model 4) gives improvements of AIC and BIC for both Enterococcus and Salmonella (Table 2). The parameters κ, v,

Table 3: Estimation results for Enterococcus. Standard deviations are given in parenthesis and bold face numbers indicate siginificant estimates.

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5
			Initialisation			
C_1	**3.06**	**3.32**	**3.32**	**3.29**	**3.30**	**3.30**
	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
C_2	**0.41**	**0.54**	**0.54**	**0.61**	**0.58**	**0.58**
	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.02)
			Drift parameters			
μ	**0.22**					
	(0.01)					
η	**0.94**	**0.94**	**0.94**	**0.93**	**0.93**	**0.93**
	(0.004)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
β		**0.12**	**0.12**	**0.36**	**0.19**	**0.18**
		(0.06)	(0.06)	(0.1)	(0.05)	(0.04)
v		**0.25**	275	**0.28**	5.9	**0.35**
		(0.01)	(320)	(0.01)	(19.4)	(0.02)
κ			915		16.3	
			(1070)		(55.3)	
			Diffusion parameters			
σ_B	**0.05**	**0.04**	**0.04**	**0.04**	**0.04**	**0.04**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
γ	**20.97**	**40.3**	**40.3**	**22.0**	31.5	30.5
	(6.52)	(12.8)	(18.6)	(10.7)	(17.0)	(15.8)
$\log(\sigma_E)$		**−1.15**	**−1.15**	**0.39**	0.11	0.09
		(0.16)	(0.16)	(0.17)	(0.16)	(0.15)
			Variance of observation noise			
$\log(s_y^2)$	**−10.18**	**−10.25**	**−10.25**	**−10.37**	**−10.32**	**−10.32**
	(0.15)	(0.15)	(0.15)	(0.14)	(0.14)	(0.14)

γ and $\log(\sigma_E)$ are not significant (Tables 3 and 4). That $\log(\sigma_E)$ is insignificant does not imply that $\sigma_E = 0$, but rather that $\sigma_E = 1$, which is not a reasonable alternative hypothesis. The alternative hypothesis for γ is $\gamma = 1$, which has been tested with resulting a large increase in AIC and BIC, for both Salmonella and Enterococcus. The large and insignificant values of κ and v imply that $\beta + v \simeq v$ and the consequence is a decoupling of v and β in the enzyme equation, while there is no real Monod term in the equation for bacteria levels. This implies that a simplified drift term not containing β in the first term might be appropriate, i.e. an enzyme process given by

$$dE_t = (v\lambda_t - E_t^2 v\lambda_t - \beta E_t)dt + \sigma_E E_t(1 - E_t)dw_E, \tag{36}$$

Table 4: Estimation results for Salmonella. Standard deviations are given in parenthesis and bold face numbers indicate siginificant estimates.

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5
			Initialisation			
C_1	**1.93**	**2.11**	**2.11**	**2.14**	**2.15**	**2.15**
	(0.09)	(0.09)	(0.09)	(0.09)	(0.08)	(0.08)
C_2	**0.12**	**0.16**	**0.16**	**0.19**	**0.26**	**0.26**
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
			Drift parameters			
μ	**0.22**					
	(0.01)					
η	**1.08**	**1.09**	**1.09**	**1.09**	**1.08**	**1.09**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.01)	(0.002)
β		**0.18**	**0.18**	**0.40**	**0.16**	**0.16**
		(0.06)	(0.06)	(0.07)	(0.03)	(0.03)
v		**0.26**	308	**0.26**	266	**0.3**
		(0.01)	(570)	(0.01)	(200)	(0.01)
κ			914		885	
			(1690)		(670)	
			Diffusion parameters			
σ_B	**0.07**	**0.07**	**0.07**	**0.07**	**0.07**	**0.07**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
γ	**68.9**	**78.4**	**78.7**	**81.5**	**59.4**	**78.1**
	(9.9)	(7.43)	(10.8)	(10.9)	(7.04)	(8.85)
$\log(\sigma_E)$		−6.11	−6.10	**0.64**	0.31	0.30
		(9.48)	(62.8)	(0.11)	(0.19)	(0.16)
			Variance of observation noise			
$\log(s_y^2)$	**−9.65**	**−9.72**	**−9.71**	**−9.74**	**−9.87**	**−9.87**
	(0.13)	(0.13)	(0.14)	(0.14)	(0.14)	(0.12)

with $\lambda_t = S_t$ and the bacteria process as given in (35). This model is referred to as Model 5 in the estimation tables (Tables 3 and 4). AIC and BIC improved for both Enterococcus and Salmonella and based on the information criteria the best choice is Model 5.

In general parameters with the same impact across models $(C_1, C_2, \eta, \gamma, \sigma_B, s_y^2)$ are relatively constant (with the exception of C_2 for salmonella) across models. This imply that estimated starting concentration (C_1 and C_2), onset of the transition to the stationary phase (η), random behaviour of the stationary phase (γ and σ_B) and observation noise (s_y^2) are constant across models. While the (random) bacteria drift term, determined by β, v, κ and σ_E varies across models.

Comparing autocorrelation functions of the residuals show large improvements for Enterococcus when comparing Model 0 and the more complex model (Figure 5), while the improvement between Model 1 and 5 is small (even though evident). These improvements are not evident for Salmonella, one reason might be that the assumption of constant variance of the standardised residuals does not apply. For high starting concentrations the variance increases with time while the variance decreases for low starting concentrations, and the results do not imply a simple modification of the data transformation.

As suggested by the likelihood results the differences between Model 4 (not shown) and 5 are not visible in an autocorrelation plot.

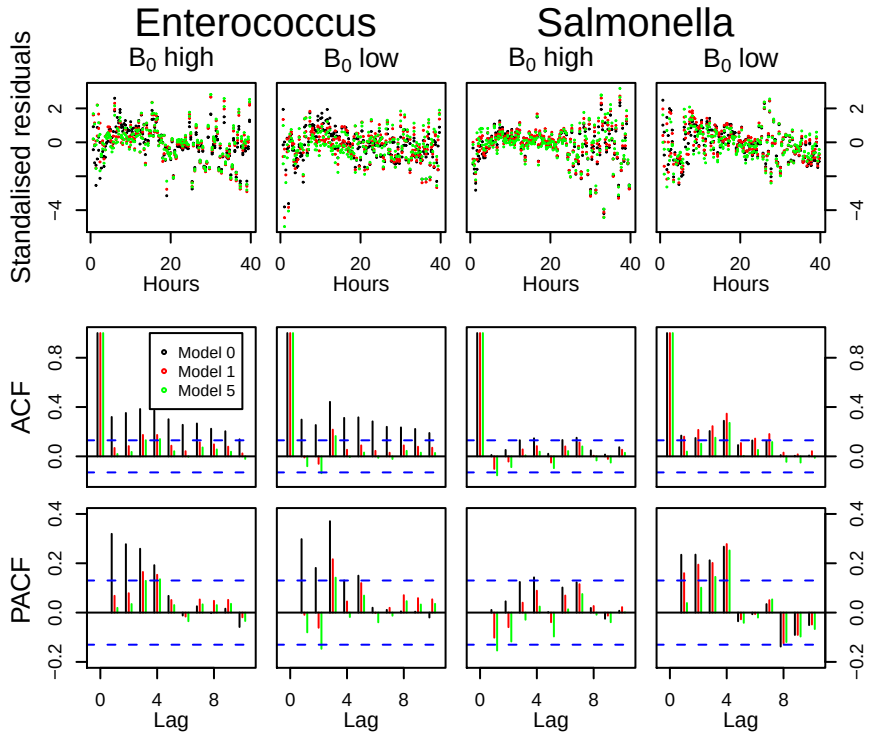

Figure 5: Standardised residuals form Model 0, 1, and 5 (top row). Bottom rows show autocorrelation functions and partial autocorrelation functions for residuals of each of the models.

8 Simulation study

The purpose of this section is to analyse the distributional behaviour of the best model from Section 7 (Model 5). The optimisation in Section 7 is based on maximum likelihood estimation, which corresponds to one-step predictions (in this case 20 minutes), and therefore a comparison between data and simulated distribution will give information on our ability to predict the distribution of a future experiment. Simulations are performed with an Euler scheme (Kloeden & Platen, 1999) on the Lamperti transformed process.

Since the scope of this section is a discussion of the distributional properties of the models, and not model selection, we will only use visual inspection. It should be noted that quantitative methods for comparing the distributional properties of models like quantile skill scores exist (e.g. Gneiting & Raftery (2007)).

Since the observation noise is small compared to the diffusion of the models, the smoothen state and data will be close and rather than comparing data and estimated distribution, we have chosen to compare the smoothen state and the distribution. A further advantage of this approach is that the smoothen state exist for both bacteria and enzyme concentration.

For the bacteria levels (based on Salmonella data) the smoothen state and the mode of the unconditional distribution differs substantially for high starting concentrations, while the performance is better for low starting concentrations (Figure 6, top row). Such behaviour is clearly undesirable and indicate that the model description is not fully adequate. Obviously one feature that is missing is the ability to describe the transition to the stationary phase. Data shows a small decline in bacteria level after the maximum have been reached, while the model can only predict positive average growth rates. The consequence is that the expected value of bacteria level from the model will converge to 1 (measure in substrate units). This is seen as the state, where all substrate is converted to bacteria acting as an effectively absorbing state.

For the smoothen enzyme levels we see a very fast increase form zero to values close to one and then a rapid decline to values close to zero when the maximum level of bacteria levels is reached (Figure 6, bottom row). The smoothen state is close to the mode of the unconditional distribution. The behaviour of the presented paths are, however qualitatively very different from the smoothen path of enzyme levels. The smoothen path starts by a rapid increase followed by a constant level close to one and then a rapid decrease, while the simulated path switches between high and low enzyme levels at high frequencies. This behaviour is due to the high estimated value of the diffusion parameter.

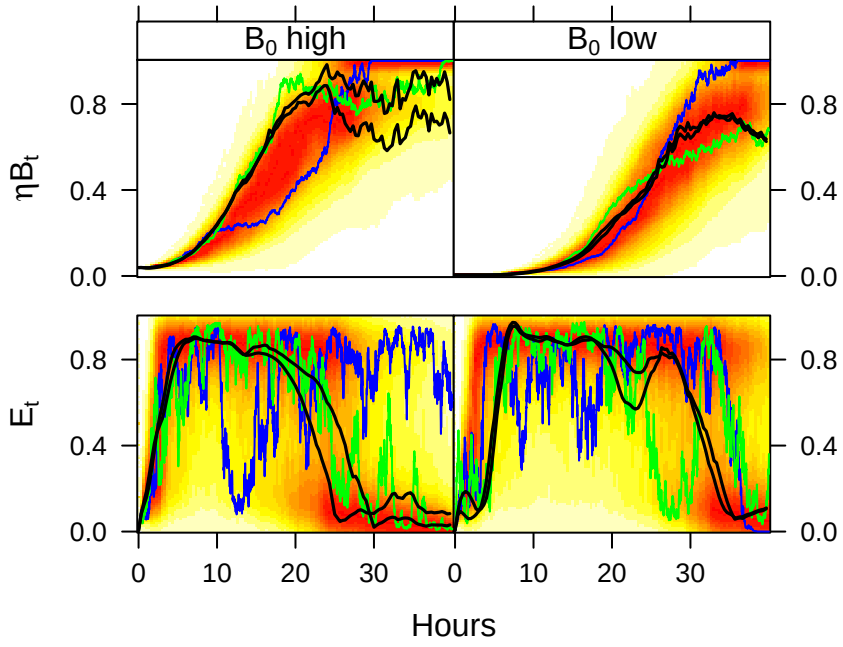

Figure 6: Smoothed states (black lines), simulations (blue and green lines) for Model 5 estimated with the Salmonella data, background colour indicate the density of simulated data. Left column is high start concentrations and right column is low start concentrations. Parameters are given in Table 4. The density estimates are based on 1000 independent realizations of the process, simulated with the Euler scheme (on the transformed process) with $\Delta t = 1s$.

9 Discussion

This paper presents an analysis of two sets of bacteria growth data in a SDE-setting; traditionally such data are analysed in a deterministic setting. The SDE formulation includes a logistic diffusion term and we argue that this inclusion is the most simple diffusion term that obey the natural constraints of the system. By introducing the power γ the shape of the diffusion function can be controlled. The large estimated values of γ indicate that the diffusion is proportional to the state, but state proportional diffusion is not reasonable since such models can produce negative substrate concentrations.

Inclusion of a random walk growth rate indicated how the model could be improved by the inclusion of an additional state. The need for an extra state was supported by the analysis of the autocorrelation of the standardised residuals from a simple model with growth rate proportional to the available substrate. The resulting model, which includes an enzyme catalysing growth was analysed with different enzyme diffusion terms. This analysis stressed that the formulation of enzyme diffusion is important for the conclusions, we are able to draw from the analysis. While state proportional enzyme diffusion did not support Monod growth, the conclusion is that the Monod term should be included if the diffusion was included as a logistic diffusion term. The inclusion of the logistic noise term did however not support the Monod effect, but rather a decoupling of parameters. In general the analysis has emphasised the importance of a proper inclusion of the diffusion term.

The set up analysed here did not include bacteria diffusion in the enzyme process, such an approach gives a simpler and more flexible (regarding Lamperti transforms) set up, it should be noted that the inclusion of bacteria diffusion in the enzyme process has been tested (with enzyme diffusion proportional to enzyme concentration). As it gave similar results it was chosen to include the simpler set up only, but the derivation of the system equation of the Lamperti transform is given in Appendix A.

By comparing the smoothen state (equivalent to data) and estimated densities, it was reviled that the models are not very well suited for simulation studies and that further model development might be appropriate. The models do lack a mechanistic description of the transitions to the stationary phase, and one possibility would be to exclude some of the data after this transition from the analysis. In this case we would however be faced with the problem of deciding a proper truncation point. Another possibility is the inclusion of a mortality parameter which seems to reasonable by looking at data, but since knowledge of the mechanistic behaviour in the stationary phase is not available this is not a trivial task.

Further it has been considered to include more parameters in the random walk analysis to obtain data driven hypothesis on the time development of the parameters. As the main scope of this article is a proper introduction of the diffusion term we have chosen to leave these analysis for future studies.

References

Baadsgaard M., Nielsen J. N., Spliid H., Madsen H., and Preisel M. (1997) Estimation in stochastic differential equations with state dependent diffusion term. *SYSID '97 - 11th IFAC symposium on system identification, IFAC*

Bajpai-Dikshit J., Suresh A.K., and Venkatesh K.V. (1997) An optimal model for representing the kinetics of growth and product formation by lactobacillus rhamnosus on multiple substrates. *Journal of Biosceince and Bioengineering*, **96**, (5), 481-486

Doshi, P., Rengaswamy, R., and Venkatesh, K. V., 1997. Modelling of microbial growth for sequential utilization in a multisubstrate environment. *Process Biochemistry* **32** (8), 643–650.

Feller W. (1951) Two singular diffusion problems, *The Annals of Mathematics, Second Series.* **54**(1), 173-182

Gard T.C. (1988) Introduction to stochastic differential equations. *Marcel Dekker, Inc.*, New York.

Gneiting T., and Raftery A.E. (2007) Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, **102**(477), 359-378

Iacus S.M. (2008) Simulation and Inference for Stochastic Differential Equations - With R Examples. *Springer Series in Statistics*

Jazwinski A.H. (1970) Stochastic Processes and Filtering Theory. *Academic Press, New York, ASU*

Kloden P., and Platen E (1999) Numerical solutions of Stochastic Differential Equations. *Springer-Verlag*

Kompala, D. S., Ramkrishna, D., and Tsao, G. T., (1984). Cybernetic modelling of microbial growth on multiple substrates. *Biotechnology and Bioengineering* **26** (11), 1272–1281.

Kovárová-Kovar, K., and Egli, T., (1998). Growth kinetics of suspended microbial cells: From single-substrate-controlled growth to mixed-substrate kinetics. *Microbiology and Molecular Biology Reviews* **62** (3), 646–666.

Kristensen N.R., and Madsen H. (2003) Continuous time stochastic modelling - CTSM 2.3 - Mathematics Guide, *Technical University of Denmark*

Kristensen N.R., Madsen H., and Jørgensen S.B. (2004a) Parameter estimation in stochastic grey-box models. *Automatica*, **40**, 225-237

Kristensen N.R., Madsen H., and Jørgensen S.B. (2004b) A method for systematic improvement of stochastic grey-box models. *Computers & Chemical Engineering*, **28**, 1431-1449

Liu, Y. (2006). A simple thermodynamic approach for derivation of a general Monod equation for microbial growth. *Biochemical Engineering Journal* **31**, 102–105.

Lobry, J., Flandrois, J., Carret, G., and Pave, A., (1992). Monod's bacterial growth model revisited. *Bulletin of Mathematical Biology* 54, 117–122.

Luschgy, H., and Pagés, G. (2006) Functional quantization of a class of Brownian diffusions: A constructive approach. *Stochastic Processes and their Applications* **116**, *310-336*

Madsen, H., Holst, J., and Thyregod, P. (1987) A Continuous Time Model for the Variations of Air Temperature. *10. Conference on Probability and Statistics in Atmospheric Science, American Meteorological Society*, 52-58, Edmonton.

Madsen H. (2008) *Time Series Analasis.* Chapman & Hall/CRC

Monod, J. (1949). The growth of bacterial cultures. *Annual Review of Microbiology* **3**, 371–394.

Patnaik, P. R., (1999). Transient sensitivity analysis of a cybernetic model of microbial growth on two substrates. *Bioprocess Engineering* 21 (2), 135–140.

Pedersen M.W., Righton D., Thygesen U.H., Andersen K., and Madsen, H. (2008) Geolocation of North Sea cod using Hidden Markov Models and behavioral switching. *Canadian Journal of Fisheries and Aquatic Sciences*, **65**, 2367-2377

Philipsen K.R., Christiansen L.E., Hasman H., and Madsen H. (2010a) Modelling conjegation with stochastic differential equations. *Journal of Theoritical Biology*, **263**, 134-142

Philipsen K.R., Christiansen L.E., Mandsberg L.F., Hasman H., and Madsen H. (2010b). Comparison of calibration curves for the relation between optical density and viable count bacteria data. *Submitted.*

Schurz H. (2007). Modelling, analysis and discretization of stochastic logistic equations. *International journal of numerical analysis and modelling*, **4**(2), 178-197

Tornøe C.W., Jacobsen J., Pedersen O., Hansen T., and Madsen H. (2004) Grey-box Modelling of Pharmacokinetic/Pharmacodynamic Systems. *J. Pharmacokinet. Pharmacodyn*, **31**, 401-417.

Øksendal B. (2003). Stochastic Differential Equations - An Introduction with Applications, Sixth edition. *Springer*, Berlin

A Including bacteria diffusion in the enzyme process

The SDE extension of Enzyme level is given by

$$dE_t = (v+\beta)\lambda_t dt - E_t d\log(B_t) - \beta E_t dt + \widetilde{\sigma}_E(t,\omega_t)dw_2 \tag{A.1}$$

in (34) bacteria levels are only included by the deterministic formulation, but by Itôs lemma we can also include bacteria diffusion.

By Itôs lemma we get

$$d\log(B_t) = (Ev\lambda_t - \frac{1}{2}(T_0 - \eta B_t)^2)dt + \sigma_B(T_0 - (\eta B)^\gamma)dw_1 \tag{A.2}$$

inserting in (A.1) gives

$$\begin{aligned} dE_t =& ((v+\beta)\lambda_t - \beta E_t)dt - E_t(E_t v\lambda_t - \frac{1}{2}(T_0 - (\eta B_t)^\gamma)^2)dt + \\ & E_t\sigma_B(T_0 - \eta B_t)dw_1 + \widetilde{\sigma}_E(t,\omega_t)dw_2 \qquad (A.3) \\ =& \left((v+\beta - vE_t^2)\lambda_t - E_t\left(\beta - \frac{1}{2}(T_0 - (\eta B_t)^\gamma)^2\right)\right)dt - \\ & \sigma_B E_t(T_0 - (\eta B_t)^\gamma)dw_1 + \widetilde{\sigma}_E(t,\omega_t)dw_2 \qquad (A.4) \end{aligned}$$

Now choose the transformation

$$Z_t^2 = \log(B_t) + \log(E_t) \tag{A.5}$$

then

$$\begin{aligned} dZ_t^2 =& \frac{dB_t}{B_t} + \frac{dE_t}{E_t} - \frac{1}{2}\left(\frac{(dB_t)^2}{B_t^2} + \frac{(dE_t)^2}{E_t^2}\right) \qquad (A.6) \\ =& (E_t v\lambda_t - \frac{1}{2}\sigma_B^2(T_0 - (\eta B_t)^\gamma)^2)dt + \sigma_B(T0 - (\eta B_t)^\gamma)dw_1 + \\ & \left(\left(\frac{v+\beta}{E_t} - vE_t\right)\lambda_t - \beta + \frac{1}{2}\sigma_B^2(T_0 - (\eta B_t)^\gamma)^2\right)dt - \\ & \sigma_B(T0 - (\eta B_t)^\gamma)dw_1 + \frac{\widetilde{\sigma}_E(t,\omega_t)}{E_t}dw_2 \\ & - \frac{1}{2}\left(\sigma_B^2(T_0 - (\eta B_t)^\gamma)^2 + \frac{\widetilde{\sigma}_E^2(t,\omega_t)}{E_t^2}\right)dt \qquad (A.7) \\ =& \left(\frac{v+\beta}{E_t}\lambda_t - \beta - \frac{1}{2}\left(\sigma_B^2(T_0 - (\eta B_t)^\gamma)^2 + \frac{\widetilde{\sigma}_E^2(t,\omega_t)}{E_t^2}\right)\right)dt + \\ & \frac{\widetilde{\sigma}_E(t,\omega_t)}{E_t}dw_2. \qquad (A.8) \end{aligned}$$

If $\widetilde{\sigma}_E(t,\omega_t)$ is chosen as proportional to E_t then dZ_2 has constant diffusion and estimation is available through CTSM. Now we have already seen that a better approach is to choose $\widetilde{\sigma}_E(t,\omega_t)$ as the logistic diffusion. Unfortunately the Lamperti transform with an explicit inverse cannot be derived in this case, it is however possible for $\widetilde{\sigma}_E(t,\omega_t) = E_t f(z_2)$ where f is a function that is simple enough to allow an explicit inverse of the Lamperti transform. Choose e.g.

$$\widetilde{\sigma}_E(t,\omega_t) = E_t(T_0 - \eta e^{Z_t^2}). \tag{A.9}$$

The state space of E_t now depend on B_t. and actually E_t is only restricted to $(0,1)$ when $\eta B_t = 1$.

In this case choose the transformation

$$Z^3 = \psi^3(Z^2) = \frac{1}{T_0}\log\left(\frac{e^{Z^2}}{T_0 - \eta e^{Z^2}}\right), \tag{A.10}$$

with the inverse given by

$$Z^2 = \log\left(\frac{T_0 e^{T_0 Z^3}}{1 + \eta e^{T_0 Z^3}}\right). \tag{A.11}$$

The derivatives of ψ^3 are given by

$$\psi^3_{z_2} = \frac{1}{T_0 - \eta e^{z_2}} = \frac{1}{T_0 - \eta E_t B_t} \tag{A.12}$$

$$\psi^3_{z_2 z_2} = \frac{\eta e^{z_2}}{(T_0 - \eta e^{z_2})^2} = \frac{\eta E_t B_t}{(T_0 - \eta E_t B_t)^2} \tag{A.13}$$

and

$$\begin{aligned} dZ_t^3 =& \left\{\frac{v+\beta}{E_t}\lambda_t - \beta - \frac{1}{2}\left(\sigma_B^2(T_0 - (\eta B_t)^\gamma)^2 + \sigma_E^2\frac{E_t^2(T_0 - \eta B_t E_t)^2}{E_t^2}\right)\right\}\times \\ & \frac{dt}{T_0 - \eta E_t B_t} + \sigma_E dw_2 + \frac{1}{2}\frac{\eta E_t B_t}{(T_0 - \eta B_t E_t)^2}\sigma_E^2\frac{E_t^2(T_0 - \eta E_t B_t)^2}{E_t^2}dt \qquad \text{(A.14)} \\ =& \left\{\frac{1}{T_0 - \eta E_t B_t}\left(\frac{v+\beta}{E_t}\lambda_t - \beta - \frac{1}{2}\sigma_B^2(T_0 - (\eta B_t)^\gamma)^2\right) - \right. \\ & \left.\frac{1}{2}\sigma_E^2(T_0 - \eta B_t E_t)\right\}dt + \sigma_E dw_2 + \frac{1}{2}\eta\sigma_E^2 E_t B_t dt \qquad \text{(A.15)} \\ =& \left\{\frac{1}{T_0 - \eta E_t B_t}\left(\frac{v+\beta}{E_t}\lambda_t - \beta - \frac{1}{2}\sigma_B^2(T_0 - (\eta B_t)^\gamma)^2\right) + \right. \\ & \left.\frac{1}{2}\sigma_E^2(2\eta B_t E_t - T_0)\right\}dt + \sigma_E dw_2 \qquad \text{(A.16)} \end{aligned}$$

with the inverse

$$E_t = \frac{e^{z_t^2}}{B} \tag{A.17}$$

$$= e^{z_t^2} \frac{\eta}{T_0} \frac{(1 + e^{\gamma T_0 Z_t^B})^{1/\gamma}}{e^{T_0 Z_t^B}} \tag{A.18}$$

$$= \frac{\eta}{T_0} \frac{T_0 e^{T_0 Z_t^3}}{1 + \eta e^{T_0 Z_t^3}} \frac{(1 + e^{\gamma T_0 Z_T^B})^{1/\gamma}}{e^{T_0 Z_t^B}} \tag{A.19}$$

$$= \eta \frac{e^{T_0 Z_t^3}}{1 + \eta e^{T_0 Z_t^3}} \frac{(1 + e^{\gamma T_0 Z_t^B})^{1/\gamma}}{e^{T_0 Z_t^B}}. \tag{A.20}$$

Actually the transformed process does not have a reasonable asymptotic behaviour unless some restrictions are imposed on the relation between β and v. Also an attempt to estimate through CTSM (with no restriction on the relation between β and v), lead to a break down of the estimation procedure illustrating the importance of considering the asymptotic behaviour.

A.1 Required relation between β and v

We consider the case where $\lambda_t = S_t$, further we set $T_0 = 1$, the critical point is $\eta E_t B_t = 1 \Rightarrow E_t = \frac{1}{\eta B_t}$, if this should work for arbitrary σ_B, we get the restriction

$$\frac{v + \beta}{E_t}(1 - \eta B_t) - \beta < 0, \tag{A.21}$$

inserting the critical point $\left(E_t = \frac{1}{\eta B_t}\right)$ gives

$$\eta B_t (v + \beta)(1 - \eta B_t) - \beta < 0, \tag{A.22}$$

the left had side is a quadratic polynomial in ηB_t with its maximum at $\eta B_t = \frac{1}{2}$, and we get the restriction

$$\frac{v + \beta}{2E_t} - \beta < 0, \tag{A.23}$$

and the critical point is $E_t \to 2$ and we get

$$\frac{v + \beta}{4} - \beta < 0 \quad \Rightarrow \quad \beta > \frac{1}{3} v. \tag{A.24}$$

So in this case the required asymptotic behaviour is ensured when $\beta > \frac{1}{3} v$, of course the appropriateness of the model has not been discussed, and we have not attempted in incorporate these restriction in the model.